



UNIVERSIDAD  
**COMPLUTENSE**  
MADRID

**Proyecto de Innovación**  
**“Innova Docencia”**

**Convocatoria 2018/2019**  
**Proyecto número 275**

**R-adaptación de la asignatura de métodos  
econométricos en economía y finanzas del Grado de  
Estadística Aplicada**

Responsable del Proyecto  
**Lorenzo Escot**

**Facultad de Estudios Estadísticos**  
**Departamento de Economía Aplicada, Pública y Política**



U N I V E R S I D A D  
**COMPLUTENSE**  
M A D R I D

**Proyecto de innovación  
“Innova Docencia” de la UCM  
Convocatoria 2018-2019  
Proyecto número 275**

**R-adaptación de la asignatura de métodos econométricos en economía y finanzas del  
Grado de Estadística Aplicada**

---

**Lorenzo Escot<sup>1</sup>, Alicia Pérez-Alonso<sup>1</sup>, José Luis Brita-Paja<sup>2</sup>, Magdalena Ferrán-Aranaz<sup>2</sup>,  
José A. Fernández-Cornejo<sup>3</sup>, Julio E. Sandubete<sup>1</sup>**

<sup>1</sup>*Dpto. Economía Aplicada, Pública y Política,  
Facultad de Estudios Estadísticos, Universidad Complutense de Madrid*

<sup>2</sup>*Dpto. Estadística y Ciencia de los Datos,  
Facultad de Estudios Estadísticos, Universidad Complutense de Madrid*

<sup>3</sup>*Dpto. Economía Aplicada, Pública y Política,  
Facultad de Ciencias Económicas y Empresariales, Universidad Complutense de Madrid*

*RESUMEN: El objetivo fundamental de este proyecto es el de adaptar todo el contenido práctico de la asignatura de Métodos Econométricos de Economía y Finanzas de cuarto curso del Grado de Estadística Aplicada de la Facultad de Estudios Estadísticos al software libre R (dejando de usar el software comercial Eviews). Para ello se han preparado una serie de materiales y recursos didácticos que se han puesto a disposición de los alumnos en el Campus Virtual. También hemos integrado en el curso el uso de la plataforma de autoaprendizaje de R "DataCamp for classroom", y se ha preparado diverso material específico de autoevaluación en el aula mediante las plataforma "Kahood".*

## Contenido

1. Objetivos propuestos en la presentación del proyecto .....	1
2. Objetivos alcanzados .....	3
3. Metodología empleada en el proyecto .....	5
4. Recursos humanos.....	7
5. Desarrollo de las actividades .....	8
6. Conclusiones y futuras líneas .....	16
7. Referencias Bibliográficas .....	17
8. ANEXOS.....	19
<i>Práctica 0: Breve introducción a R .....</i>	<i>20</i>
<i>Práctica 1: Lectura de datos desde la web del INE: La identificación del ciclo económico y la ley de Okun.....</i>	<i>41</i>
<i>Práctica 2 : La Ley de Okun: ¿a qué ritmo debe crecer la Economía para comenzar a generar empleo .....</i>	<i>117</i>
<i>Práctica 3 : La Ley de Okun: Test de Cambio Estructural, ¿han sido “eficaces” las reformas del mercado de trabajo? .....</i>	<i>140</i>
<i>Práctica 4 : Ley de Okun con datos de la UE desde la web de Eurostat .....</i>	<i>169</i>
<i>Práctica 5 : La ecuación de salarios y la discriminación por razón de género .....</i>	<i>195</i>
<i>Práctica 6 : Detección y estimación robusta ante heterocedasticidad: Un ejemplo sencillo con el gasto medio por persona en las CC.AA. ....</i>	<i>224</i>
<i>Práctica 7 : Heterocedasticidad: La demanda de Cigarrillos ¿Qué política antitabaco es más eficiente? .....</i>	<i>251</i>
<i>Práctica 8 : Abracadabra. No estacionariedad en media y Cointegración .....</i>	<i>268</i>
<i>Práctica 9 : Cointegración y Modelos de Corrección del Error .....</i>	<i>295</i>
<i>Práctica 10: De vueltas con la ley de Okun: cointegración de las series de PIB y Empleo.....</i>	<i>312</i>
<i>Práctica 11: Autocorrelación en La Ley de Okun.....</i>	<i>354</i>
<i>Práctica 12: El Índice de Confianza del Consumidor como indicador adelantado del PIB.....</i>	<i>382</i>
<i>Práctica 13: Análisis univariante de la serie del PIB y ARIMAX con tasas de ocupados .....</i>	<i>409</i>
<i>Práctica 14 : Endogeneidad con un ejemplo sencillo: Estimación y Diagnósis.....</i>	<i>426</i>
<i>Práctica 15 : Estimación del rendimiento de la educación para las mujeres casadas .....</i>	<i>438</i>
<i>Práctica 16: Modelo de datos de Panel con R: la librería “pml” .....</i>	<i>457</i>
<i>Práctica 17: La ecuación de Salarios con Datos de Panel.....</i>	<i>493</i>
<i>Práctica 18: Regresión cuantílica para la discriminación salarial y distribución de salarios .....</i>	<i>534</i>
<i>Práctica 19 : Modelos de Econometría Espacial con R: La ley de Okun espacial.....</i>	<i>558</i>
<i>Ejemplos de la plataforma Kahoot.....</i>	<i>595</i>

## 1. Objetivos propuestos en la presentación del proyecto

Con la aparición de los nuevos planes de estudio adaptados al Espacio Europeo de Educación Superior, la Universidad Complutense de Madrid (UCM) realizó una profunda revisión de los contenidos formativos y de la metodología docente de la antigua Diplomatura de Estadística que quedó sustituida por la oferta de un nuevo Grado de Estadística Aplicada en la Facultad de Estudios Estadísticos. En este contexto de adaptación, se incorporó en el cuarto curso del nuevo grado la asignatura de *Métodos Econométricos en Economía y Finanzas*. Con esta asignatura se pretendía complementar las destrezas del alumno de Estadística Aplicada ofreciéndole la posibilidad de adquirir el conocimiento de un conjunto de métodos de regresión ampliamente utilizadas en los estudios econométricos, técnicas y herramientas estadísticas no estudiadas en el resto de asignaturas cursadas en el grado, o utilizadas con una diferente aplicación, identificación aquéllas más adecuada para el análisis cuantitativo de la Economía atendiendo en cada caso a las características de los fenómenos objeto de estudio, los datos disponibles y las propiedades de las variables explicativas y explicadas analizadas.

El contenido de esta asignatura se presenta al alumno desde un enfoque eminentemente práctico: primero se proporcionan los principales elementos teóricos de cada capítulo (supuestos, contrastes, métodos de estimación y diagnosis), para pasar, posteriormente, a la aplicación práctica de cada uno de esos elementos teóricos.

Las aplicaciones prácticas a través de ejemplos y ejercicios con ordenador se han venido impartiendo con el software *Eviews*, que tiene como principal ventaja que se adapta muy bien al contenido de la asignatura (la ayuda de este software es un auténtico y excelente manual de Econometría). Sin embargo, al tratarse de un software comercial, su principal desventaja es el coste, tanto para la UCM como para los propios estudiantes. Asimismo, la evidencia de los últimos años muestra una tendencia a la baja en cuanto a popularidad y uso del *Eviews*, que no aparece en los rankings del software estadístico más utilizado por parte de las empresas potencialmente empleadoras de nuestros egresados en Estadística Aplicada. Por el contrario, entre el software estadístico más demandado por parte de estas empresas se encuentran *Python* y *R*, en código abierto, y *SPSS*, *SAS*, *Stata* y *MATLAB*, como programas comerciales (Muenchen, 2017). Esto supone una clara desventaja para nuestros alumnos.

En la enseñanza de la Econometría, *Gretl* es un software econométrico de código abierto muy popular entre los docentes, que permite realizar en clase los ejemplos y ejercicios de los principales manuales de la materia. Sin embargo, no permite realizar análisis y técnicas de regresión más avanzadas o especializadas, y precisamente por eso ofrece un interfaz que da acceso al paquete estadístico de software libre *R*. Es por ello que, tampoco hemos optado por esta opción de *Gretl*, habiendo fijado como objetivo fundamental en nuestro caso pasar del uso del software comercial *Eviews* al software libre *R* ([r-project.org](http://r-project.org)). Este cambio se sustenta en tres motivos fundamentales:

- *R* es un software de libre distribución, lo que facilita el acceso de todos los estudiantes a esta herramienta. Esto puede suponer un ahorro de recursos tanto para los alumnos como para la propia UCM.



- Existe una extensa comunidad de usuarios de R, que proporcionan no sólo ayuda con el uso de este software estadístico, sino que realiza aportaciones en forma de librerías, paquetes o funciones con los últimos desarrollos técnicos y los nuevos métodos econométricos que van apareciendo en la literatura. Esto hace que R (y también python, en alguna medida también STATA) cuente con el conjunto de técnicas más avanzado que otros paquetes (como Gretel, Eviews, o SAS, que se demoran más tiempo en incorporar estas técnicas en sus respectivos paquetes).
- Precisamente, las aportaciones de la comunidad de usuarios en forma de funciones y librerías a disposición de todos los usuarios, permiten que todas las técnicas y contrastes que se incluyen en la asignatura están disponibles en R, lo que permite ahorrar tiempo en programación que puede dedicarse a la aplicación práctica de los contenidos.
- El uso de R (y otros como SAS o Python) en la docencia es una demanda creciente de los alumnos de Estadística Aplicada que responde, como ya se ha dicho, al hecho de ser software libre y al uso cada vez más generalizado de este lenguaje en su futuro profesional.

Frente a estas ventajas, R también tiene sus inconvenientes. Precisamente el principal objetivo de este proyecto de innovación docente es tratar de minimizar dichos inconvenientes y limitaciones de R, aportando recursos didácticos para las clases en el aula, para el autoaprendizaje y la autoevaluación del alumno adaptados al contenido de la asignatura.

**Objetivo 1** Adaptar el **contenido práctico de la asignatura** de Métodos Econométricos en Economía y Finanzas del Grado de Estadística Aplicada al software estadístico *R*:

- 1.1 Preparación de un **índice de las librerías y funciones estadísticas** con las que poder aplicar las técnicas de análisis econométrico que se ven de forma teórica en la asignatura;
- 1.2 Además de proporcionar al alumno el repertorio de librerías que podría utilizar, se requiere mostrar al alumno cómo **implementar en la práctica dichas funciones y librerías**, siguiendo el procedimiento y etapas de análisis según se presentan en las clases teóricas.
- 1.3 En la medida en que *R* es un lenguaje vivo en el que continuamente se ponen a disposición de toda la comunidad nuevos contenidos, también es necesario proporcionar a los estudiantes herramientas para que sean autónomos y ellos mismos **aprendan a aprender R**; a la hora de buscar nuevas funciones y librerías, así como solucionar dudas y problemas que le puedan sobrevenir
- 1.4 En la medida de que uno de los objetivos de la asignatura es también la identificación y presentación de fuentes de información económica, se propone como objetivo mostrar cómo desde el propio *R* puede realizarse la consulta, descarga y análisis de los datos proporcionados por las principales fuentes oficiales de estadísticas públicas como el INE, Eurostat, el Banco de España, o el Ministerio de Economía.

**Objetivo 2** Proporcionar a los alumnos el apoyo de material didáctico sobre el uso general de *R* para el análisis de datos. En este sentido, y de forma complementaria a ese material didáctico preparado específicamente para la asignatura, pretendemos integrar los recursos disponibles en la plataforma externa de autoaprendizaje *online* del software estadístico *R* de *Datacamp* en la plataforma *Moodle* de la UCM, CVUCM-Moodle 3.4. En particular, esto se hace a través de la aplicación *DataCamp For The Classroom*, que permite asignar tareas a los alumnos dentro de la plataforma e incorporarlas en el campus virtual.

**Objetivo 3** Preparar materiales y recursos didácticos para el autoaprendizaje de cada capítulo, tanto de la teoría, sobre las diferentes técnicas de análisis y diagnosis, como del

software *R*. De esta forma, las clases presenciales pueden dedicarse fundamentalmente a realizar actividades de aprendizaje para la asimilación de los contenidos de la asignatura.

**Objetivo 4** Aplicar técnicas de aprendizaje como la Gamificación y otras metodologías activas en Educación. En particular, el objetivo es implementar dos recursos para la autoevaluación de los alumnos. Uno de ellos presencial, en el aula, utilizando la aplicación *Kahoot*, que puede servir como instrumento para reforzar los contenidos más importantes de cada capítulo; y otro no presencial, implementando la librería *R/exams* para la autoevaluación *online* de los alumnos.

## 2. Objetivos alcanzados

Como ya hemos comentado anteriormente, el uso de *R* tiene muchas ventajas frente a otros programas comerciales. Pero también presenta desventajas o limitaciones que con este proyecto pretendemos reducir o mitigar. La principal desventaja, desde nuestro punto de vista, del uso de *R* en nuestra asignatura, está en la alta disponibilidad de librerías y funciones que ofrece, ya que en muchos casos, existe más de una (y más de dos) funciones que realizan exactamente lo mismo pero que están programadas y pertenecen a diferentes librerías. Como ya hemos comentado, con *R* es posible implementar todos los métodos y técnicas que se tratan en la asignatura, pero no todas esas técnicas y contrastes están disponibles en un paquete único de econometría que se ajuste al contenido de la asignatura (existen intentos de unificación en *R* como *R-comander* o *Rattle*, pero que tampoco se ajustan a los contenidos de nuestra asignatura). En este sentido **el método de aprendizaje con *R* es el contrario al utilizado en paquetes comerciales como *SPSS* o *Eviews***, donde sin conocer todas las técnicas, se puede ir explorando por los menús y por sus documentos de ayuda para ver que técnicas ofrece el programa e investigar cada una de ellas. Por el contrario, con el uso de *R*, el método de utilización (y aprendizaje) es justo el contrario. No existe un menú de técnicas disponibles, ni un manual de ayuda, por lo que ***es necesario saber a priori que técnica se quiere aplicar y entonces buscar la librería o librerías que lo hacen posible***.

En este sentido, la *R*-adaptación de la asignatura ha requerido proporcionar al alumno de ese repertorio de técnicas a utilizar (en forma de paquetes, librerías y funciones) y el conocimiento de cómo detectar la idoneidad de su uso en cada aplicación práctica. En este sentido, se ha realizado un repertorio de scripts o prácticas en *R* en el que no sólo se muestran el índice de librerías y funciones que mejor se adaptan a los procedimientos, métodos y contrastes de cada capítulo de la asignatura, sino que en esos scripts se presenta también la implementación de cada una de las etapas (contrastos, estimación de modelos, diagnosis, predicción,...) que desde el punto de vista teórico se presentan previamente al alumno en cada uno de los capítulos.

Podemos afirmar, por tanto que en general se han alcanzado todos los objetivos propuestos:

1. Se han preparado un conjunto de 19 scripts de *R* en el que se presenta la implementación en *R* de cada uno de los métodos propuestos en cada capítulo de la asignatura (Test de cambio estructural, Heterocedasticidad; Autocorrelación; Raíces Unitarias y Cointegración, Endogeneidad y Mínimos Cuadrados en dos etapas; Datos de panel; Regresión Cuantílica; y modelos de Econometría Espacial) dando así cumplimiento a los dos objetivos fundamentales del PIE (**objetivo 1 y objetivo 3**)

(véase más adelante el detalle de cada uno de estos Capítulos y los correspondientes scripts en *R* preparados que se ofrecen como ANEXOS a esta memoria). Estos scripts se han ofrecido a los alumnos antes de clase a través del Campus Virtual y se han ido ejecutando en clase en los laboratorios de informática. De esta forma, los alumnos podían ir resolviendo en la propia clase todas las dudas y preguntas que pudiesen surgir al respecto. Cabe mencionar que esta adaptación de la asignatura a *R* es la que más recursos, tiempo y trabajo de preparación previa a requerido al grupo de trabajo del presente proyecto de innovación educativa

2. Para cubrir la necesidad de presentar a los alumnos cómo desde el propio *R* pueden *descargarse datos del INE, de Eurostat o del Banco de España*, todas las prácticas utilizan datos descargados directamente desde estas instituciones, por lo que los scripts incorporan las funciones y métodos necesarios no sólo para estimar los diferentes modelos de regresión, sino también para realizar dicha descarga de datos, su depuración y análisis descriptivo inicial (**objetivo 1**).
3. Como ya se ha comentado, el problema con los repertorios de librerías y funciones en *R* es su obsolescencia. En efecto, *R* es un lenguaje vivo y continuamente están apareciendo nuevas librerías que incorporan nuevas funciones, por lo que no sólo hay que proporcionar a los alumnos listados o repertorios de funciones y librerías que pueden utilizar. También es necesario proporcionarles las herramientas para que ellos mismos "**aprendan a aprender *R***" (**objetivo 1**). Es por ello que a lo largo de las diferentes clases se les ha mostrado a los alumnos dónde buscar las funciones que necesitan en cada caso; o como pedir ayuda a la comunidad *R* para encontrar la solución a problemas específicos que en el futuro les puedan aparecer. En este sentido se han presentado a los alumnos:
  - 3.1 Las [CRAN Task View de \*R\*](#), que proporciona un listado de librerías existentes, clasificado por temas y actualizado por el propio equipo de *r-project*,
  - 3.2 Los blogs o foros de consulta como [r-bloggers](#), o [stack overflow](#). Nosotros hemos recurrido ampliamente a este tipo de páginas a la hora de preparar todo nuestro material para la asignatura, y hemos transmitido a los alumnos la importancia de su uso, ya que pueden ser de mucha utilidad a los alumnos en el desarrollo futuro de sus propias aplicaciones estadísticas.
4. El uso aplicado de software específico para el análisis econométrico en una asignatura de 4º curso del Grado de Estadística Aplicada, requiere un conocimiento previo de los alumnos del lenguaje *R*. En este sentido los alumnos tienen que cursar una asignatura de *Software Estadístico* científico donde desde hace ya varios años se presenta el programa *R*. Sin embargo no todos los alumnos han aplicado *R* en otras asignaturas antes de llegar a la nuestra. Por ello consideramos necesario disponer de recursos para suplir las carencias o deficiencias en el uso de *R* de algunos de nuestros alumnos. En efecto, nuestra experiencia reciente (que esperamos que cambie pronto en el futuro) nos indica que algunos alumnos llegan a nuestra asignatura con un nivel muy básico de *R* (como decimos, por lo menos, en los últimos dos cursos académicos). En este sentido creemos que es muy importante reforzar el conocimiento general del lenguaje *R* de nuestros alumnos. Y por ello la adaptación de la asignatura al uso de *R* ha requerido la preparación de material didáctico que permita la **adquisición por parte de nuestros alumnos, de una forma autónoma, de las destrezas básicas sobre el paquete *R***. En este sentido:

- 4.1 Se han preparado diferentes materiales didácticos de “**Introducción a R**”, de “**análisis exploratorio de datos con R**” y sobre cómo “**programar funciones propias en R**” (**objetivo 3**). En este punto ha sido fundamental el carácter interdepartamental del equipo de este proyecto, que cuenta con dos profesores del departamento de Estadística y Ciencia de los datos, que han ayudado en la creación de estos recursos.
- 4.2 De forma complementaria a estos materiales, consideramos importante poner a disposición de nuestros alumnos otros recursos necesarios para reforzar su autoaprendizaje online del lenguaje R. Existen muchos portales MOOC (Coursera, Edx, MiriadaX, etc ) que ofrecen cursos de iniciación a R. Entre ellos, consideramos a los cursos ofrecidos por [DataCamp](#) como una de las opciones más atractivas. Esta plataforma *DataCamp* ofrece una gran variedad de cursos y recursos didácticos para el aprendizaje de lenguajes R y Python. Así, hemos abierto nuestros propios grupos dentro de la aplicación ***DataCamp for The Classroom***, lo cual ha permitido asignar a los alumnos tareas dentro de la plataforma de *DataCamp* y lo que es más importante, les ha dado acceso ilimitado a todo el contenido, cursos y materiales considerados premium o de pago dentro de esta plataforma (**objetivo 2**). Sin embargo, en este punto no hemos conseguido, como nos habíamos propuesto inicialmente, integrar *DataCamp for the Classroom* en la plataforma Moodle del campus virtual. Técnicamente no podía realizarse, ya que recientemente *DataCamp* ha deshabilitado esta posibilidad de integrar su plataforma en *Moodle*. A pesar de ello, consideramos que el uso de *DataCamp for the classroom* ha sido una de las principales aportaciones o revelaciones de nuestro proyecto. Recomendamos ya a todos nuestros compañeros que utilicen este recurso en sus propias asignaturas y lo que nos comunican tanto ellos como nuestros alumnos es que efectivamente es una herramienta de autoaprendizaje del alumno claramente satisfactoria
5. Finalmente, y en cuanto al **objetivo 4**, de los dos recursos inicialmente propuestos para aplicar técnicas de evaluación, tanto presencial en el aula (*Kahoot*), como de autoevaluación del alumno (*R/exams*), sólo hemos podido poner en práctica la primera de ellas. La utilización de los *Kahoot* ha permitido aplicar un enfoque de gamificación que se ha valorado muy positivamente por parte de los alumnos, convirtiéndose en un instrumento muy eficaz para reforzar y repasar los contenidos más importantes de cada capítulo en el aula. Esta herramienta permite discutir y confirmar con los propios alumnos la correcta asimilación de los diferentes métodos y contenidos de la asignatura así como sus aplicaciones en R

### 3. Metodología empleada en el proyecto

El procedimiento a desarrollar y alcanzar los cuatro grandes objetivos de la R-adaptación de la asignatura puede describirse del siguiente modo.

1. Se ha revisado el programa de la asignatura para revisar el contenido del mismo y analizar la viabilidad de su aplicación práctica en R, y evaluar la cantidad de casos prácticos y ejercicios en R que podían incluirse en cada uno de los capítulos del temario

2. Dada la necesidad de refrescar (y en algún caso aislado iniciar) el uso de R, se ha preparado una documentación de iniciación con el software R. Esta documentación inicial, junto con las instrucciones básicas de cómo instalar R y RStudio, los tipos de elementos y estructura y enfoque de utilización de R se incluyó en un primer capítulo que denominamos capítulo 0, para destacar su carácter previo a los propios contenidos de la asignatura.
3. Precisamente para ayudar a los alumnos con el aprendizaje de R dimos de alta nuestros dos grupos (cuarto A – profesor Escot; y cuarto B- profesora Pérez-Alonso) en *DataCamp for the Classroom*. Estos grupos de clase son gratuitos para las universidades y dan acceso premium a todos los contenidos de *DataCamp* a los alumnos del grupo durante 6 meses. Además como profesores del grupo podemos asignarle tareas en forma de ejercicios o lecciones específicas de alguno o algunos cursos de *DataCamp* (por supuesto también cursos completos, cuya duración promedio, por cierto, suele estar en torno a las 4 horas), ya que la plataforma permite hacer seguimiento de qué alumno del curso ha cumplimentado cada uno de los curso y tareas que nosotros le asignemos dentro de *DataCamp*
4. Elaboración de recurso didáctico: Carga de datos desde la web del INE, ESUROSTAT y BANCO de ESPAÑA desde R y análisis exploratorio inicial.
5. Búsqueda de las funciones y librerías de R por parte del equipo que mejor se adaptaban al temario de la asignatura.
6. Preparar y diseñar contenidos y **aplicaciones prácticas** para la asignatura con R.
7. Elaboración de documentos en *RMarkdown* con la implementación de las aplicaciones prácticas en R para cada uno de los temas de la asignatura, esto es, ejercicios y ejemplos prácticos en R correspondientes a los métodos y contrastes específicos de cada capítulo: Modelo lineal, estimación y diagnóstico; Test de cambio estructural; Heterocedasticidad; Autocorrelación; Raíces unitarias y cointegración; Endogeneidad y mínimos cuadrados en dos etapas; Datos de panel; Regresión cuantílica; Modelos de econometría espacial. Como ya se ha comentado, para la elaboración de los ejemplos y prácticas en R se ha utilizado la aplicación *RMarkdown*, que proporciona la posibilidad de elaborar documentos en *pdf* o presentaciones *html* donde se combinen explicaciones o aclaraciones en texto formateado para una mejor comprensión de lo que el código de R trata de realizar en cada caso; junto al propio código en R con el conjunto de funciones empleadas en cada ejercicio práctico; y acompañado de forma adicional a las salidas o resultados que deberían obtenerse. La utilización de este formato *RMarkdown* resulta, a nuestro entender, más didáctico que un simple script o programa que contenga sólo el conjunto del código en R de cada práctica.
8. Elaboración de recursos para la autoevaluación en clase a través de *Kahoot*, que sirvan como instrumento para ir evaluando el seguimiento de la asignatura y para repasar y reforzar los elementos fundamentales de cada capítulo.
9. Presentación, aplicación y utilización de todos los recursos y materiales didácticos preparados en las clases presenciales. Dichos materiales se pusieron a disposición del alumnado en el campus virtual, de esta forma, una vez presentados inicialmente en clase, cada uno de ellos podía descargarlo y revisarlo de manera autónoma para preparar la asignatura, preguntando las dudas y problemas que les fueron apareciendo, a través de consultas por correo electrónico, o de forma presencial tanto en tutorías como en las clases en el aula. La necesidad de trabajar estos recursos didácticos por parte del alumnado, consideramos que ha tenido como resultado la participación activa de los alumnos en su aprendizaje de la asignatura.

## 4. Recursos humanos

Equipo del Proyecto:

1. **Lorenzo Escot Mangas**  
(PDI UCM responsable del proyecto, Dpto de Economía Aplicada, Pública y Política, Facultad de Estudios Estadísticos) escot@ucm.es
2. **Alicia Pérez Alonso**  
(PDI UCM, Dpto de Economía Aplicada, Pública y Política, Facultad de Estudios Estadísticos) apereza@ucm.es
3. **José Luis Brita Paja**  
(PDI UCM, Dpto de Estadística e Investigación Operativa III, Facultad de Estudios Estadísticos) jlbrita@estad.ucm.es
4. **Magdalena Ferrán Aranaz**  
(PDI UCM, Dpto de Estadística e Investigación Operativa III, Facultad de Estudios Estadísticos) maenafer@estad.ucm.es
5. **José Andrés Fernández Cornejo**  
(PDI UCM, Dpto de Economía Aplicada, Pública y Política, Facultad de Ciencias Económicas y Empresariales) jafercor@ccee.ucm.es
6. **Julio Emilio Sandubete Galán**  
(Personal investigador en formación contratado no doctor –becas UCM–, Dpto de Economía Aplicada, Pública y Política, Facultad de Estudios Estadísticos) jsandube@ucm.es

Si bien han sido los dos profesores directamente encargados de la docencia, el profesor Lorenzo Escot y la profesora Alicia Pérez-Alonso, son los que han tenido una mayor responsabilidad en la R-adaptación de los contenidos prácticos de la asignatura, sobre todo en lo que se refiere en su puesta en práctica con el alumnado, todos los miembros del equipo del proyecto han colaborado en la preparación de los materiales didácticos y en la preparación de las clases, considerándose que la aportación de todos ellos ha sido relevante y necesaria para el éxito del proyecto.

Consideramos en este sentido especialmente satisfactoria la participación de los profesores Jose Luis Brita y Magdalena Ferrán Aranaz del departamento de Estadística y Ciencia de los Datos. En efecto, la asignatura de métodos econométricos está a cargo del departamento de Economía Aplicada, Pública y Política, y por ello los cuatro profesores del equipo del proyecto de este departamento hemos tenido la responsabilidad de liderar y poner en práctica la R-adaptación de la asignatura. Sin embargo, la R-adaptación de nuestra asignatura requiere de nuestro alumnado algún tipo de conocimiento previo de este software estadístico. En este sentido existe en el programa actual de estudios del Grado de Estadística Aplicada una asignatura de segundo curso “Software estadístico” (impartida por el profesor José Luis Brita), en la que se ha comenzado a impartir recientemente el software R para iniciar a los alumnos en el lenguaje R. Igualmente, en otras asignaturas, como las impartidas por la profesora Ferrán Aranaz también se ha decidido comenzar a aplicar R, como lenguaje estadístico para realizar las prácticas. La participación de nuestros dos compañeros del departamento de Estadística en este proyecto consideramos que ha sido esencial para preparar recursos y materiales para la introducción genérica a R; al uso de R en el análisis exploratorio de los datos; y al uso de R en la estimación de modelos de regresión. De forma adicional, consideramos que su participación en este proyecto también ha sido relevante para que ellos puedan tener un mejor conocimiento de primera mano de nuestras necesidades desde la asignatura de econometría, esto es, saber cuáles serán las necesidades futuras de los

alumnos del grado de Estadística Aplicada cuando lleguen a nuestra asignatura, al menos en cuanto al manejo de R en el análisis multivariante de datos. Esperamos que esta participación permitirá, en la medida de lo posible, una mayor coordinación en cuanto a la docencia de software estadístico a lo largo de los diferentes cursos, y en particular entre nuestra asignatura y otras impartidas por el departamento de Ciencia de los Datos, como la asignatura de “software estadístico” (que tiene un tiempo limitado para la docencia de R), y alguna de las asignaturas de “análisis multivariante” (que además de SAS y SPSS también comienza a utilizar R). Esperamos que todo ello se traduzca en un mejor dominio de este lenguaje de programación estadística por parte de nuestros alumnos del grado, y en un mejor aprovechamiento de nuestra asignatura de métodos econométricos situada en el plan de estudios al final del grado, en el último semestre del último curso del grado de Estadística Aplicada.

## 5. Desarrollo de las actividades

En cuanto al desarrollo de las actividades, una parte importante de los recursos didácticos creados han sido los scripts de R en los que se aplicaban las técnicas y procedimientos de cada uno de los 10 capítulos. En total han sido 20 scripts (incluyendo el capítulo inicial de introducción a R). El contenido de estos scripts se adjunta a la presente memoria en forma de ANEXOS. A continuación se detallan los contenidos desarrollados, **en forma de prácticas**, según cada capítulo del programa de la asignatura:

### Capítulo 0 INTRODUCCIÓN AL MANEJO DE R

***Introducción a la plataforma DataCamp for the classroom***

***Práctica 0: Breve introducción a R***

### Capítulo 1 INTRODUCCIÓN

- Naturaleza y contenido de la Econometría
- La modelización económica
- Planteamiento del curso

***Práctica 1: Lectura de datos desde la web del INE: La identificación del ciclo económico y la ley de Okun***

### Capítulo 2 REGRESIÓN LINEAL

- El Modelo Lineal General (MLG)
- Hipótesis del modelo
- Estimación por Mínimos Cuadrados Ordinarios (MCO)
- Inferencia y predicción en el MLG
- Contrastes de cambio estructural
- Multicolinealidad y errores de medida

***Práctica 2 : La Ley de Okun: ¿a qué ritmo debe crecer la Economía para comenzar a generar empleo***

***Práctica 3 : La Ley de Okun: Test de Cambio Estructural, ¿han sido “eficaces” las reformas del mercado de trabajo?***

***Práctica 4 : Ley de Okun con datos de la UE desde la web de Eurostat***

***Práctica 5 : La ecuación de salarios y la discriminación por razón de género***



### Capítulo 3 MODELOS CON DATOS DE SECCIÓN CRUZADA I: HETEROCEDASTICIDAD

- Causas de la heterocedasticidad
- Contrastes de heterocedasticidad
- El Estimador de Mínimos Cuadrados Generalizados
- El Estimador de Mínimos Cuadrados ponderados factible

**Práctica 6 :** *Detección y estimación robusta ante heterocedasticidad: Un ejemplo sencillo con el gasto medio por persona en las CC.AA.*

**Práctica 7 :** *Heterocedasticidad: La demanda de Cigarrillos ¿Qué política antitabaco es más eficiente?*

### Capítulo 4 MODELOS CON DATOS DE SERIES TEMPORALES I: COINTEGRACIÓN Y REGRESIONES ESPURIAS

- Magia “potagia” o las regresiones espurias
- Cointegración y modelos de regresión con series no estacionarias
- ¿Cuántas veces hay que diferenciar una serie para convertirla en estacionaria? El contraste de raíces unitarias
- Contrastes de cointegración
- Cointegración y Mecanismos de Corrección del Error (MCE)

**Práctica 8 :** *Abracadabra: No estacionariedad e media y Cointegración*

**Práctica 9 :** *Cointegración y Modelos de Corrección del Error*

**Práctica 10:** *De vueltas con la ley de Okun: cointegración de las series de PIB y Empleo*

### Capítulo 5 MODELOS CON DATOS DE SERIES TEMPORALES II: Autocorrelación, indicadores adelantados y causalidad en sentido de Granger

- Modelos de regresión con series temporales estacionarias: autocorrelación
- Contrastes de autocorrelación
- El estimador de Mínimos Cuadrados Generalizados (MCG)

**Práctica 11:** *Autocorrelación en La Ley de Okun*

- Indicadores adelantados: Causalidad en sentido de Granger y correlación contemporánea

**Práctica 12:** *El Índice de Confianza del Consumidor como indicador adelantado del PIB*

### Capítulo 6 MODELOS CON DATOS DE SERIES TEMPORALES III: MODELOS DINAMICOS ARIMA y ARIMAX

- Series temporales estacionarias y Orden de integración I(d)
- Modelos ARIMA
- Estimación, Diagnóstico de modelos ARIMA
- Extensión: Modelos ARIMAX

**Práctica 13:** *Análisis univariante de la serie del PIB y ARIMAX con tasas de ocupados*

### Capítulo 7 ENDOGENEIDAD Y EL ESTIMADOR DE VARIABLES INSTRUMENTALES

- ¿En qué consiste el problema de la Endogeneidad?, ¿qué problemas origina?
- Orígenes de la endogeneidad:
  - Omisión de alguna variable relevante (porque no es observable) que está correlacionada con alguna otra variable explicativa
  - Errores en la medida
  - Modelo de regresión con retardos y autocorrelación en los errores



- Estimador con variables instrumentales y mínimos cuadrados en dos etapas
- Contrastes de endogeneidad y restricciones de sobreestimación
- Práctica 14 : Endogeneidad con un ejemplo sencillo: Estimación y Diagnóstico***
- Práctica 15 : Estimación del rendimiento de la educación para las mujeres casadas***

## Capítulo 8 MODELOS DE REGRESIÓN CON DATOS DE PANEL

- Especificación de modelos con datos de panel
- Estimación de modelos con efectos fijos
- Estimación de modelos con efectos aleatorios
- Sistema de ecuaciones aparentemente no relacionadas
- Estimador de efectos fijos observables con el estimador CRE
- Práctica 16: Modelo de datos de Panel con R: la librería “pml”***
- Práctica 17: La ecuación de Salarios con Datos de Panel***

## Capítulo 9 MODELOS DE REGRESIÓN CUANTÍLICA

- Regresión en media y regresión por cuantiles
- Estimación de la regresión intercuantílica
- Representación gráfica de la solución
- Práctica 18: Regresión cuantílica para la discriminación salarial y distribución de salarios***

## Capítulo 10 MODELOS DE ECONOMETRÍA ESPACIAL

- Introducción a la estadística y a las medidas de asociación espacial
- La construcción de la matriz de distancias para la medición de dependencias espaciales
- Especificaciones del modelo con dependencia espacial: modelo de error con retardos espaciales y el modelo con errores espaciales autoregresivos
- Problemas de mínimos cuadrados ordinarios para la estimación de modelos con retardo espacial y el estimador de mínimos cuadrados en dos etapas
- Práctica 19 : Modelos de Econometría Espacial con R: La ley de Okun espacial***

Estas 20 prácticas en R se han ido presentado a lo largo de los diferentes capítulos del temario de la asignatura. Los alumnos disponían a través del Campus Virtual (plataforma Moodle) antes de iniciar cada capítulo, además de material para preparar los contenidos teóricos (apuntes, presentaciones y otras referencias bibliográficas), los scripts y documentos *Rmarkdown* en R con los que se realiza cada aplicación práctica de esos contenidos teóricos. Estos scripts se han ido ejecutando en las clases presenciales, en los laboratorios de informática reservados para impartir nuestras clases. De esta forma, los alumnos podían ir resolviendo en la propia clase todas las dudas y preguntas que pudiesen surgir al respecto. Igualmente se iban proponiendo y asignando a los alumnos diferentes tareas y cursos desde la plataforma *DataCamp for the classroom* que a nuestro juicio mejor podían servir para mejorar y completar cada capítulo.

El objetivo de la oferta tanto de los recursos didácticos de *DataCamp* como de los materiales específicos desarrollados por el equipo de este proyecto de innovación educativa para nuestra asignatura persiguen el objetivo de un autoaprendizaje del alumno, de forma que se permita ir modificando el método docente, pasando la clase magistral a

un método donde exista una preparación previa de los alumnos de los contenidos de cada capítulo (con los recursos que los profesores han puesto a su disposición en el campus virtual), y donde las clases presenciales se reserven para la aplicación práctica y activa del alumno frente al ordenador de las diferentes actividades y prácticas correspondientes a cada capítulo. Sería una aproximación al concepto de aula invertida (“flipped classroom”) propuesta inicialmente por Lage, Platt y Treglia(2000) y Bergmann y Sams (2012), donde las clases presenciales se utilizan para reforzar la asimilación de los contenidos: resolución de dudas; actividades de aprendizaje con R; aprendiendo a aprender R (cómo buscar funciones y librerías necesarias para la resolución de problemas); resolución de ejercicios aplicados; desarrollo de las prácticas propuestas; y la evaluación y repaso de contenidos fundamentales de cada capítulo de la asignatura. En este enfoque docente, los materiales y recursos didácticos ofrecidos online a los alumnos, juegan un papel fundamental, y deberían ser consultados por éstos tanto antes como después de las clases presenciales para reforzar su aprendizaje.

Al finalizar cada capítulo se dedicaban 15-20 minutos para realizar una evaluación presencial de los conocimientos adquiridos mediante el uso de la plataforma *Kahoot* (se muestran también en el ANEXO algunos de los *Kahoots* utilizados). Estos “juegos”, han permitido afianzar y repasar los principales contenidos de cada capítulo con una implicación activa de por parte del alumnado. Ante la imposibilidad de desarrollar procedimientos de auto evaluación mediante aplicaciones de *R-exams* (que se plantean como objetivo a alcanzar en próximos años), al final de cada capítulo se han propuesto una serie de tareas para casa en la que los alumnos tenían que aplicar estos mismos scripts con otro conjunto de datos, o incluso con los mismos conjuntos, pero solicitando test y aplicaciones adicionales que requerían del alumno la ejecución de todo el proceso. Algunas de dichas tareas y ejercicios se han utilizado para realizar la evaluación continua del alumnado. En particular, las tareas propuestas que han servido para realizar la evaluación continua han sido las siguientes.

---

### **TAREA 1. Ley de Okun (primera parte)**

Tenéis que dar respuesta razonada a las siguientes tres preguntas:

- 1) ¿Hemos salido de la crisis económica? ¿cuándo?
- 2) ¿A qué ritmo tiene que crecer la economía española para comenzar a generar empleo?.
- 3) ¿Se van a crear 2 millones de empleos durante el 2019 y 2020 en la economía Española?, ¿no?, ¿cuantos entonces?
- 4) ¿Han sido eficaces las reformas del mercado de trabajo?

1980 Estatuto de los trabajadores;  
 1984: inicio de la temporalidad;  
 1988: Nuevos contratos Temporales y basura;  
 1992: reforma del seguro desempleo;  
 1994: reforma importante que da entrada entre otras a la ETT;  
 1997: reforma pactada;  
 2001-2002: reforma Aznar con huelga y marcha atrás;  
 2006: reforma leve; 2009: reforma leve;

2010 reforma importante ante la crisis continuada en 2012, con una Reforma importante

El trabajo se presentará en Word/pdf/html y contará al menos con los siguientes apartados:

- Introducción (¿qué se va a hacer?)
- Metodología y datos (¿cómo se va a hacer?: fundamentación en la Ley de Okun, datos, métodos de regresión por MCO. ¿qué datos se van a utilizar?)
- Resultados y diagnóstico
- Conclusiones al trabajo

Se valorará positivamente la presentación del script R utilizado

---

## **TAREA 2. ¿Están discriminadas las mujeres en nuestro mercado de trabajo?**

Tenéis que estimar el tamaño de la brecha salarial en España utilizando el fichero de salarios que utilizamos en la práctica de clase:

- 1) Utilizando una ecuación de salarios, tenéis que contrastar si a las mujeres se les remunera en media de manera diferente que a los varones
  - 2) Utilizar la metodología de Oaxaca-Blinder para estimar qué parte de la brecha salarial se debe a diferencias en características y qué parte se debe a discriminación
  - 3) Estimar cuánto tendría que aumentar el salario medio de las mujeres si fuesen remuneradas como los hombres
  - 4) Estimar cuánto tendría que bajar el salario medio de los varones si éstos fuesen remunerados como las mujeres
  - 5) Mostrar un histograma con la distribución de los indicadores de discriminación de las mujeres.
- 

## **TAREA 3 (Voluntaria): ¿Cuánto vale la vida de un individuo?**

Se trata de una práctica voluntaria (no es obligatoria para aprobar el curso por evaluación continua). Sí que será necesaria si queréis mejorar vuestra nota final por evaluación continua.

Tenéis que responder a tres preguntas

1. ¿Cuánto vale la vida de un individuo?
2. ¿Vale la pena estudiar un Grado Universitario?
3. ¿Qué vale más la vida de un hombre o la de una mujer?
4. ¿Y cuánto más vale la vida de un hombre con estudios superiores frente a los hombres con estudios secundarios? ¿y en las mujeres, cuánto más ganan las que tienen estudios superiores?

Para contestar a estas preguntas hay que calcular el valor actual esperado del flujo de rentas que obtendrá un individuo medio, comparar el valor actual esperado de alguien que tiene estudios de secundaria y compararlo con el que decide seguir estudiando, o comparar el valor actual de un hombre con el de una mujer, y lo mismo pero teniendo en cuenta diferencias de sexo y niveles de estudio

Para ello podéis hacer los siguientes supuestos.

- 0) Considerar en todos los casos la vida de un individuo de 18 años de edad y utilizar los datos que ya empleasteis para hacer la práctica 2 de discriminación salarial para estimar la ecuación de Salarios considerando que dicha ecuación corresponde al año en que el individuo tiene los 18 años.
  - 1) Todos los años posteriores los salarios se indician con la inflación
  - 2) La inflación se mantendrá constante en el futuro con un valor del 1%
  - 3) Los tipos de interés se mantendrán constantes en el futuro con un valor del 2%
  - 4) Los individuos se jubilan a los 67 años con una pensión que se calcula con la media del salario percibido por el trabajador en los últimos 20 años y que en todo caso no puede superar los 2000 euros en el momento de la jubilación. Una vez calculada la jubilación esa pensión se indicia anualmente con la tasa de inflación.
  - 5) Las tasas de mortalidad se publican en el INE
- 

#### **TAREA 4 (Voluntaria): ¿Cuánto vale mi casa?**

Esta es una segunda práctica Voluntaria que sirve para subir nota en la evaluación continua (no es obligatoria para aprobar pero sí para obtener buena nota)

El objetivo es saber cuánto podrías pedir por el alquiler/venta de vuestra casa. Para ello utilizareis la fórmula de la rentabilidad del alquiler y la estimación del precio de venta de vuestra vivienda.

- 1) Para calcular la rentabilidad del alquiler en vuestra zona comparareis el precio medio de venta y de alquiler según la página web de idealista.com:  $\text{rentabilidad} = (\text{alquiler mensual} \times 12) / \text{precio de venta}$
- 2) Para conocer el precio de venta (el precio total, no el precio en euros por m<sup>2</sup>) hay que construir un modelo de regresión de precios hedónicos para determinar el precio de la vivienda en función de sus características (metros cuadrados, número de habitaciones, número de baños, plazas de garaje, planta, orientación, piscina, aire acondicionado, calefacción central, etc.). Los datos hay que obtenerlos de internet del portal idealista.com y buscar los 40 testigos/anuncios más cercanos a vuestra casa (mínimo 30), y estimar un modelo de regresión.

Antes de realizar la regresión tenéis que hacer un breve análisis descriptivo de las variables que habéis seleccionado, incluyendo un histograma del precio total de la vivienda, que será vuestra variable explicada.

Estaréis haciendo un modelo con datos de sección cruzada, por lo que tenéis que guardar una especial atención a la posible presencia de heteroscedasticidad en vuestra regresión.

Una vez calculado el modelo tendréis que predecir el precio teórico de venta de vuestra casa sustituyendo en la regresión estimada las características de vuestra propia vivienda y obteniendo su precio ajustado.

- 3) Con este precio y la rentabilidad del alquiler de vuestra zona podéis estimar cual será el precio al que en teoría podríais alquilar vuestra vivienda aplicando la fórmula de la vivienda como bien de inversión (utilizando la rentabilidad estimada en 1) y el precio de venta en 2) sólo tenéis que despejar el precio mensual del alquiler de la fórmula).

---

#### **TAREA 5. ¿Están Cointegradas las series de PIB y EMPLEO?**

Tenéis que utilizar las series de PIB (índice de Volumen encadenado) y Empleo (puestos de trabajo equivalente a tiempo completo) para intentar averiguar si estas series están cointegradas.

En caso afirmativo intentad estimar un modelo de corrección del error.

---

#### **TAREA 6 (Voluntaria) Indicadores adelantados del PIB**

En esta práctica hay que utilizar los indicadores adelantados del PIB para intentar mejorar las predicciones respecto a un modelo ARIMA.

- 1) ¿Son buenos indicadores adelantados del PIB El índice de Confianza del Consumidor (elaborado por la Comisión Europea) y el Indicador Sintético de la Actividad Económica (elaborado por el Ministerio de Economía)?. Tenéis que describir brevemente cada uno de estos índices y justificar si son o no son buenos indicadores adelantados (gráficamente, con correlaciones cruzadas, y con el test de Granger).
  - 2) Tomando la serie del PIB hasta el cuarto trimestre del 2018 (pero reservando los cuatro trimestres del 2018 para comprobar lo bien o mal que haceis las predicciones) construir tres modelos para hacer predicción:
    - 2.1) Modelo ARIMA
    - 2.2) Modelo de regresión con el Indicador de Confianza del Consumidor
    - 2.3) Modelo de regresión con el indicador Sintético de la Actividad Económica
  - 3) Hacer predicciones de la tasa de crecimiento del PIB del 2018 y calcular el Error Cuadrático Medios de las predicciones en cada modelo ¿con cuál de los tres predice mejor el PIB?
-

## **TAREA 7. Datos de Panel: ¿Son eficientes las políticas de subvenciones para mejorar la productividad?**

Se trata de que evaluéis la eficacia de las políticas de concesión de ayudas directas a empresas para mejorar la productividad con un modelo de regresión utilizando datos de panel. Como mínimo hay que utilizar el siguiente modelo

$$\text{Scrap}(ij) = c + c(i) + b_1 \cdot D88 + b_2 \cdot D89 + b_3 \cdot \text{Grant}(ij)$$

La variable *Scrap* es la variable que se utiliza para medir la productividad y recoge la tasa de productos defectuosos (en porcentaje). *D88* y *D89* son variables dicotómicas temporales. y *Grant* es una variable dicotómica que toma valor 1 si la empresa *i* obtuvo ayuda en el periodo *j* (0 en caso contrario). Se puede introducir alguna variable adicional, pero el interés estará siempre en el coeficiente de la variable *Grant*

Toda la información está en el conjunto de datos *data('jtrain')* de la librería *Wooldridge* (es un ejemplo de su libro en el capítulo de Datos de Panel. Con *?jtrain* obtendréis la ayuda con la descripción de las variables.

### Contenido de la práctica

- 1) Introducción de lo que vais a hacer y que son los datos de panel y la metodología o procedimiento a seguir
  - 2) Estimar modelo sin efectos fijos (pooling model)
  - 3) Estimar el modelo con el Estimador de Efectos Fijos (por si existe heteroscedasticidad utilizar estimadores robustos de la varianza, y utilizar mínimos cuadrados ponderados por los cross section)
  - 4) Contrastar la hipótesis de existencia de efectos Fijos Heterogeneos ( $c(ij)=0$  para todo *i*)
  - 5) Estimar el modelo con el Estimador de Efectos Aleatorios (por si existe heteroscedasticidad en los *cross section* utilizar la aproximación de white para estimar consistentemente las varianzas de los estimadores)
  - 6) Contraste de Hausman para saber si los efectos fijos  $c(ij)$  están relacionados con las variables explicativas (*GRANT*) y darle una explicación razonable en sentido económico
  - 7) Determinar el mejor modelo a utilizar: En caso de que sea el modelo de efectos Fijos aplicar Mínimos Cuadrados Generalizados con ponderaciones Cross-section
  - 8) Diagnósis del modelo: "...entonces, ¿son las políticas de subvenciones efectivas o no?"
- 

## **TAREA 8. Correlación Espacial de la Tasa de Paro Municipal**

En esta práctica tenéis que responder a los siguientes puntos relacionados con los modelos de regresión Espacial

- 1) Representa un mapa de España con la distribución Municipal de la Tasa de Paro, ¿Existe alguna zona de España donde la tasa de paro sea significativamente mayor? ¿por qué eso puede ser indicio de que existe dependencia espacial en la tasa de paro?
  - 2) Calcula la matriz de ponderaciones espaciales utilizando contigüidad de un nivel tipo "Reina" y construye la variable de retardo espacial de la Tasa de Paro, es decir, para cada municipio tienes que calcular la tasa media de paro de los municipios vecinos colindantes.
  - 3) ¿Existe asociación espacial en la tasa de paro municipal?. Utiliza para responder a esta pregunta el mapa de la distribución municipal del retardo espacial de la Tasa de Paro calculado en el punto anterior y la I de Moran ¿resulta significativa?, ¿cómo debe interpretarse la representación gráfica de la I de Moran?.
  - 4) Construye un modelo para explicar la tasa de paro Municipal en función de la Renta per cápita Municipal, ¿existe menos paro en los Municipios más ricos?. Analiza los residuos de este modelo inicial estimado por mínimos cuadrados, ¿presentan correlación espacial?. En caso afirmativo hay que ampliar el modelo inicial para recoger la correlación espacial en la tasa de paro Municipal. Utiliza el modelo de regresión espacial con retardo espacial en la variable dependiente "*Spatial Lag*", ¿has conseguido recoger la dependencia espacial con este modelo con retardo espacial de la variable dependiente?. Justifica
- 

## 6. Conclusiones y futuras líneas

Una vez finalizado este proyecto de R-adaptación de la asignatura de métodos econométricos en Economía y finanzas, e implementado por primera vez en el curso 2018-2019, consideramos que sus principales objetivos relacionados con el uso aplicado y práctico de software específico R para el análisis econométrico se han cumplido satisfactoriamente. De hecho tanto las encuestas realizadas por la Facultad de Estudios Estadísticos como las encuestas de evaluación del programa Docencia de la Universidad Complutense de Madrid muestran la buena aceptación por parte de los alumnos de esta R-adaptación de la asignatura. Igualmente hemos tenido la posibilidad de presentar las principales líneas del actual proyecto en dos congresos (uno nacional y otro internacional) sobre innovación en docencia, siendo muy positivos todos los comentarios recibidos:

- Ponencia: "*R-adaptation of the course "Econometric Methods in Economics and Finance"*", presentada en the **11th International Conference on Education and New Learning Technologies, EDULEARN 19**, organizada por the International Academy of Technology, Education and Development IATED, y celebrada entre el 1 y el 3 de julio de 2019, Palma de Mallorca, Spain. Publicado en EDULEARN19 Proceedings ISSN: 2340-1117, doi.org/10.21125/edulearn.2019, Pages: 1318-1326
- Ponencia: "*R-adaptación de la asignatura de Métodos Econométricos en Economía y Finanzas*", presentada en la sesión paralela 1ª de experiencias de innovación docente en las **XV Jornadas sobre Docencia de Economía Aplicada**, organizada por la

Asociación Libre de Economía, celebrada el 8 de Marzo de 2019, CaixaForum Madrid, Paseo del Prado 3, Madrid.

Los objetivos concretos planteados para este proyecto se han ceñido a la elaboración de los recursos y materiales didácticos comentados anteriormente en la memoria cuyos destinatarios finales son nuestros alumnos del Grado de Estadística Aplicada.

Sin embargo, nuestro compromiso no se limita al curso 2018-2019. Esperamos poder dar continuidad al proyecto en años sucesivos mejorando la oferta de recursos didácticos. Sin duda nos queda pendiente la integración de estas aplicaciones prácticas con la elaboración de nuevos recursos que integren estos ejercicios con la teoría, los test, métodos de regresión y diagnosis de cada tema. Nos planteamos la posibilidad de elaborar un auténtico manual teórico-práctico de la asignatura adaptada a R que podría al final plasmarse en algún tipo de librería R con las funciones que hemos ido recopilando y elaborando para la asignatura de análisis econométrico. También queda pendiente la implementación plena de los recursos elaborados con R-Exams. Incluso preparar algún tipo de curso en Español dentro de la plataforma *DataCamp* sobre econometría aplicada en R.

## 7. Referencias Bibliográficas

- Bergmann, J. & Sams, A., (2012). Flip Your Classroom. Reach Every Student in Every Class Every Day. ISTE
- Dalgaard, P. (2008): Introductory statistics with R. Springer Science & Business Media <https://link-springer-com.bucm.idm.oclc.org/book/10.1007%2F978-0-387-79054-1>
- Grün B, Zeileis A (2009). "Automatic Generation of Exams in R." Journal of Statistical Software, 29(10), 1-14. doi:10.18637/jss.v029.i10
- Hanck, Ch; Arnold, M; Gerber, A; and Schmelzer.M (2018) Introduction to Econometrics with R <https://www.econometrics-with-r.org/>
- Heiss, F. (2016): Using R for Introductory Econometrics. Create Space <http://www.urfie.net/read.html>
- Kleibei, C. & Zeileis, A. (2008): Applied Econometrics with R. Springer <https://link-springer-com.bucm.idm.oclc.org/book/10.1007%2F978-0-387-77318-6>
- Lage, M.J., Platt, G. J. & Treglia, M., (2000). Inverting the classroom: a gateway to creating an inclusive learning environment. Journal of Economic Education, 31(1), p. 30-43.
- Muenchen, R. A. (2017). The Popularity of Data Science Software. Recuperado de <http://r4stats.com/articles/popularity/>
- Pfaff, B. (2008): Analysis of Integrated and Cointegrated Time Series with R. Springer-Verlag <https://link-springer-com.bucm.idm.oclc.org/book/10.1007%2F978-0-387-75967-8>
- Shumway, R. H. & Stoffer, D. S. (2017): Time Series Analysis and Its Applications: With R Examples. 4th Edition, Springer
- Venables, W.N, Smith, D.M. & the R core Team (2018): An Introduction to R. <https://cran.r-project.org/doc/manuals/R-intro.pdf>
- Zeileis A, Umlauf N, Leisch F (2014). "Flexible Generation of E-Learning Exams in R: Moodle Quizzes, OLAT Assessments, and Beyond." Journal of Statistical Software, 58(1), 1-36. doi:10.18637/jss.v058.i01



- Zeileis A (2016a). Blended Learning Tools for Large Statistics and Mathematics Courses. Presented at DAGStat 2016, Universität Göttingen, Germany.
- Zeileis A (2016b). A One-for-All Exams Generator: Written Exams, Online Tests, and Live Quizzes with R. Presented at Psychoco 2016 - International Workshop on Psychometric Computing, Université de Liège, Belgium.
- Zuur, A., Ieno, E. N. & Meesters, E. (2009): A Beginner's Guide to R. Springer Science & Business Media [https://cds.cern.ch/record/1401934/files/9780387938363\\_TOC.pdf](https://cds.cern.ch/record/1401934/files/9780387938363_TOC.pdf)

## **8. ANEXOS**



Proyecto de innovación  
"Innova Docencia" de la UCM  
Convocatoria 2018-2019  
Proyecto número 275

**R-adaptación de la asignatura de métodos econométricos en economía y finanzas del  
Grado de Estadística Aplicada**

Facultad de Estudios Estadísticos  
Departamento de Economía Aplicada, Pública y Política

**Título**

***Práctica 0: Breve introducción a R***

**Correspondiente al**

**Capítulo 0 INTRODUCCIÓN AL MANEJO DE R**

***Práctica 0: Breve introducción a R***

# Breve introducción a R

© 2019 *Métodos Econométricos en Economía y Finanzas*

*Escot Mangas, Lorenzo & Pérez Alonso, Alicia & Sandubete, Julio E.*

*Departamento de Economía Aplicada, Pública y Política*

*Facultad de Estudios Estadísticos, Universidad Complutense de Madrid*

*04 febrero 2019*

## Descarga e instalación de R y RStudio

- **R**: <https://www.r-project.org/> (<https://www.r-project.org/>)
- **RStudio**: <https://www.rstudio.com/> (<https://www.rstudio.com/>)

Este documento está escrito en **R Markdown** (<http://rmarkdown.rstudio.com> (<http://rmarkdown.rstudio.com>)). Esta aplicación de **RStudio** es una sintaxis de formato simple para la creación de documentos y presentaciones en HTML, PDF y MS Word.

## Cómo ejecutar un programa y guardar los resultados

Podemos usar **R** de modo interactivo: introducimos una orden en la consola y **R** nos responde en la línea siguiente; pero cuando cerramos el programa no queda ninguna constancia de nuestra actividad. Sin embargo, **R** nos ofrece la posibilidad de construir *scripts* (guiones), que consisten simplemente en archivos de instrucciones/órdenes en los que se recoge toda nuestra actividad: qué datos leemos, qué transformaciones les aplicamos, qué medidas descriptivas calculamos, qué gráficos hacemos, qué análisis estadístico desarrollamos, etc.

Los programas de **R** tienen extensión `.R`. Cuando guardamos el *script*, guardamos sólo las instrucciones, no los resultados. Si queremos volver a ejecutar el programa para recuperar los resultados, tenemos que cargarlo en la memoria de **R** utilizando la función `source()`

```
source("myprogram.R")
```

Por defecto, todos los resultados aparecen en la consola de **R**. Existe, sin embargo, la posibilidad de enviar los resultados directamente a un fichero externo usando el comando `sink()`

```
args(sink)
```

```
## function (file = NULL, append = FALSE, type = c("output", "message"),  
##       split = FALSE)  
## NULL
```

Valores predeterminados de los argumentos de esta función:

- `append=FALSE` : sobrescribimos el fichero en lugar de agregar los nuevos resultados a continuación de los ya guardados.
- `split=FALSE` : los resultados se envían directamente al fichero de salida y no se imprimen en pantalla. Si `TRUE`, se envían a ambos.

Veamos un ejemplo:

```
sink("myouput.txt") #Redirige los resultados al fichero de salida
x<-2+2
x
sink() #Esta orden devuelve la salida de nuevo a la pantalla.
```

Mientras en la pantalla aparecen solamente los comandos, el fichero *myouput.txt* contiene el valor del objeto *x*. A partir de la instrucción `sink()`, los resultados aparecen de nuevo en la consola de **R**.

Si ejecutamos un comando para dibujar un gráfico, se abre una ventana gráfica en **R**, pero también se pueden dibujar gráficos directamente en un archivo externo sin tener que abrir ninguna ventana gráfica. El comando depende del tipo de archivo: `pdf("mygraph.pdf")`, `win.metafile("mygraph.wmf")`

```
* png("mygraph.png")
* jpeg("mygraph.jpg")
* bmp("mygraph.bmp")
* postscript("mygraph.ps")
```

En este caso, hay que especificar el nombre del archivo a crear como opción de la función y para cerrarlo hay que ejecutar `dev.off()`.

Veamos un ejemplo:

```
jpeg("myplot.jpg")
plot(x) #instrucción genérica para obtener un gráfico
dev.off()
```

## Entorno de trabajo (*workspace*)

Podemos visualizar/cambiar el directorio de trabajo a través de la ventana inferior derecha de **RStudio** que muestra el directorio de trabajo (pestañas *Files > More: go to working directory/set as working directory*). Si no especificamos un directorio, **R** asume que nuestro archivo está en el directorio de trabajo actual. Si queremos realizar esta acción a través de comandos de **R** ejecutamos

```
getwd()
setwd("C:/Users/romin/apereza/_myR")
getwd()
```

El comando `ls()` visualiza todos los objetos que tenemos en el directorio de trabajo.

```
ls()
```

```
## character(0)
```

Si quiero eliminar alguno de estos objetos, utilizaré el comando `rm()` especificando el nombre del objeto a eliminar. `rm(list=ls())` limpia la memoria de todos los objetos que estén en ella.

Para visualizar todas las opciones predeterminadas que usa **R** ejecutamos

```
options()
```

Si, por ejemplo, queremos visualizar el número de dígitos que utiliza **R** en su forma predeterminada (*default value*)

```
options("digits")
```

```
## $digits
## [1] 7
```

Si queremos reducir el número predeterminado de dígitos de 7 a 3

```
options(digits=3)
options("digits")
```

```
## $digits
## [1] 3
```

En ocasiones puede que queramos cambiar algunas opciones sólo de modo temporal o en una parte concreta del programa. En ese caso, antes de hacer los cambios, primero guardamos las opciones actuales para después poder restablecerlas

```
defaultopts <- options()
options(echo=T)
#Escribo mi programa aquí
options(defaultopts)
```

## Cómo obtener ayuda

El menú general de opciones de ayuda `help.start()` nos da acceso a manuales y guías de referencia. Si lo que queremos es buscar en los manuales y librerías de **R** una palabra o un concepto específico usamos

```
help.search("covariance")
??covariance
```

También podemos buscar en la web de **R** si existe una función que implemente, por ejemplo, el filtro de Kalman. Esta búsqueda nos devuelve como resultado una página web.

```
RSiteSearch("Kalman")
```

Si estamos interesados en acceder a la ayuda de un comando (función) específico

```
help(rnorm)
?rnorm
```

Si lo que queremos es visualizar los argumentos o el código de ese comando, usaremos respectivamente

```
args(rnorm)
```

```
## function (n, mean = 0, sd = 1)
## NULL
```

```
rnorm
```

```
## function (n, mean = 0, sd = 1)
## .Call(C_rnorm, n, mean, sd)
## <bytecode: 0x000000001bc55008>
## <environment: namespace:stats>
```

Para visualizar un listado de todas las funciones que contienen ese comando

```
apropos("rnorm")
```

```
## [1] "rnorm"
```

Para reproducir un ejemplo que utilice el comando que usamos

```
example(rnorm)
```

Equivalentemente, la ventana inferior derecha de **RStudio** contiene una pestaña de ayuda (*Help*). Además, en el menú principal, el icono de *Help* nos da acceso al chuletero *Cheatsheet*.

También podemos encontrar ayuda en foros de programadores como Stack Overflow (<https://es.stackoverflow.com/questions/tagged/r>). El enlace se corresponde al foro en español, aunque el foro en inglés es más amplio. El éxito en estos foros se mide en términos de que alguien responda a tu pregunta, así que es imprescindible aprender a formular preguntas de modo eficiente. De hecho, tienen una guía (<https://es.stackoverflow.com/help/how-to-ask>) sobre cómo se hace una buena pregunta.

## Librerías

Cuando instalamos **R** se incorporan por defecto numerosos paquetes de base. Sin embargo, muchas de las funciones no de base que utilizamos de **R** se almacenan en librerías específicas, así que es frecuente que tengamos que instalar un paquete antes de empezar a trabajar. En la web de **R** podemos consultar una lista (<https://cran.r-project.org/web/packages/>) de todos los paquetes disponibles. Podemos ver una lista de los paquetes instalados en nuestro ordenador ejecutando

```
library()
```

Para instalar un paquete utilizamos el comando `install.packages("nombrepaquete")`. También podemos usar la ventana inferior derecha de **RStudio** que muestra los paquetes para instalarlos directamente (pestaña *Packages*). Por el contrario, `remove.packages("nombrepaquete")` desinstala el paquete del disco duro de nuestro ordenador.

La función `old.packages()` compara los paquetes que tenemos instalados en nuestro ordenador con los que se encuentran en CRAN, y nos proporciona una lista de aquellos que cuentan con una versión más moderna. Para actualizarlos basta con ejecutar `update.packages()`.

Es importante distinguir entre tener un paquete instalado en el ordenador y tenerlo cargado en memoria. Podemos ver una lista de los paquetes cargados en la memoria ejecutando

```
search()
```

```
## [1] ".GlobalEnv"      "package:stats"    "package:graphics"
## [4] "package:grDevices" "package:utils"    "package:datasets"
## [7] "package:methods"  "AutoLoads"        "package:base"
```

Una vez descargado e instalado el paquete (¡¡esto sólo es necesario hacerlo una vez!!), para utilizar las funciones que contiene debemos cargarlo en memoria mediante

```
library(foreign)
```

La librería `foreign` contiene las funciones necesarias para leer y escribir datos que tienen formatos de otros programas como **SAS**, **SPSS** o **Stata**.

Para saber el directorio en el que guardamos las librerías usamos

```
.libPaths()
```

```
## [1] "C:/Users/romin/Documents/R/win-library/3.4"
## [2] "C:/Program Files/R/R-3.4.3/library"
```

Otro paquete muy útil es `datasets`, el cual contiene todas las bases de datos integradas en **R**. El comando `data()` nos dará una lista de los datos disponibles en el sistema base. Para visualizar una lista con los nombres utilizamos el comando `ls`

```
ls("package:datasets")[1:4]
```

```
## [1] "ability.cov"      "airmiles"         "AirPassengers"    "airquality"
```

Si lo que queremos es una lista de todas las bases de datos disponibles en todos los paquetes de **R**, no sólo las integradas en `datasets`, utilizamos `data(package = .packages(all.available = TRUE))`.

Por el contrario, si buscamos los datos de una librería específica, por ejemplo, `MASS` (*Modern Applied Statistics with S*), primero tenemos que cargar dicha librería y después podemos visualizar una lista con los nombres de las bases de datos incluidas en ese paquete.

```
library(MASS)
ls("package:MASS")[1:6]
```

```
## [1] "abbey"           "accdeaths"       "addterm"         "Aids2"           "Animals"         "anorexia"
```

Después, para cargar en la memoria una base de datos de esa librería, por ejemplo `Aids2`, que contiene los datos de supervivencia con VIH en Australia, usamos

```
data(Aids2)
```

## Tipos de datos

Todos los elementos que maneja **R** son objetos: un valor numérico es un objeto, un vector es un objeto, una función es un objeto, una base de datos es un objeto, un gráfico es un objeto, una tabla es un objeto... Cada *clase* de objeto tiene diferentes atributos que determinan la forma en que trabajan dentro de **R**.

**R** es un lenguaje orientado a objetos, lo que hace que diversas “funciones genéricas” se comporten de manera distinta según el objeto al que se aplican; pudiendo incluso no hacer nada (o producir un error) si se le pasan argumentos de una clase inadecuada. Ejemplo de ello son las funciones `summary()`, `print()`, `plot()` y `mean()` que devuelven una información distinta según el objeto al que sean aplicadas. Para visualizar el tipo de objetos a los que podemos aplicar estas funciones:

```
methods(mean)
```

```
## [1] mean.Date      mean.default    mean.difftime   mean.POSIXct    mean.POSIXlt
## see '?methods' for accessing help and source code
```

```
methods(plot)[1:3]
```

```
## [1] "plot.acf"           "plot.correspondence" "plot.data.frame"
```

```
methods(summary)[1:3]
```

```
## [1] "summary.aov"        "summary.aovlist"    "summary.aspell"
```

```
methods(print)[1:3]
```

```
## [1] "print.abbrev" "print.acf"         "print.AES"
```



**R** admite como tipo de datos básicos valores numéricos, caracteres y valores booleanos. Los distintos tipos de datos pueden almacenarse en variables que se identifican con un nombre. Los nombres de variables deben comenzar por una letra seguida de más letras, dígitos o los caracteres punto `.` o subrayado `_`. Se diferencia entre mayúsculas y minúsculas. Los nombres de las variables deben ser informativos utilizando por convenio el `.` para separar distintas palabras.

Para almacenar datos en una variable se utiliza el operador de asignación `<-`. Este operador asigna un valor a un nombre ( `expresión <- valor` ). Las órdenes se separan mediante `;` o mediante un cambio de línea. El carácter `#` indica que lo escrito a continuación es un comentario. Los comentarios son de una línea.

```
y <- 5; print(y) #Imprime de manera explícita
```

```
## [1] 5
```

Si una expresión se utiliza como una orden por sí misma, su valor se imprime en pantalla y se pierde. Por ejemplo, la siguiente orden simplemente imprime en pantalla el inverso del valor anterior.

```
1/y
```

```
## [1] 0.2
```

- Tipos de datos (*clases*) básicos o atómicos
  1. *character*: carácter/alfanumérico/cadena de texto (*strings*)
  2. *numeric*: número real
  3. *integer*: número entero
  4. *complex*: número complejo
  5. *logical*: valor lógico o booleano `TRUE` (verdadero)/ `FALSE` (falso). Estos valores también se representan por `T` / `F`

En su forma predeterminada, **R** trata los números como objetos numéricos (i.e. números reales tipo *double*). Si explícitamente queremos que sea un entero, tenemos que especificar el sufijo `L` (ej.: `1L` es un objeto numérico mientras que `1L` es un entero)

```
myname <- "Paquita Salas"  
class(myname)
```

```
## [1] "character"
```

```
myheight <- 1.65  
class(myheight)
```

```
## [1] "numeric"
```

```
typeof(myheight)
```

```
## [1] "double"
```

```
myage <- 50L  
class(myage)
```

```
## [1] "integer"
```

```
actor <- TRUE
class(actor)
```

```
## [1] "logical"
```

```
mycomplex<-2+4i
class(mycomplex)
```

```
## [1] "complex"
```

`Inf` es un número especial que representa el infinito. Podemos usarlo también para cálculos ordinarios.

```
1/0
```

```
## [1] Inf
```

```
1/Inf
```

```
## [1] 0
```

El valor `NaN` representa un valor no definido (“Not a Number”). También podríamos interpretarlo como un valor faltante (*missing value*).

```
0/0
```

```
## [1] NaN
```

[COMPLETAR] - NA : “Not Available”, no disponible. - NULL : objeto “vacío” - NaN : “Not a Number”, no es un número (ej.: `log(-1)`, `sqrt(-1)`)

## Estructuras de datos

**R** utiliza diferentes estructuras de datos. La estructura más simple es el vector. Un número, por sí mismo, se considera un vector de longitud uno.

- Estructuras de datos
  - *vector*: colección de uno o más objetos del mismo tipo (ej. caracteres, números...). Esta característica hace que se los denomine vectores atómicos (*atomic vectors*).
  - *Missing value*: Normalmente en **R** se codifican como `NA`
  - *factor*: variables cualitativas que se pueden incluir en los modelos
  - *matrix*: vector de dimensión múltiple en el que las columnas deben ser de la misma clase y longitud
  - *data frame*: vector de dimensión múltiple que permite valores de clases diferentes pero de igual longitud
  - *list*: vector que admite valores de clases diferentes
  - *time series*: [COMPLETAR]
  - *POIXct, Date*: Fechas/horas [COMPLETAR]

## Vectores

### Generación de vectores

En **R** existen varias funciones para generar vectores:

- concatenación `c()` : función que combina objetos (homogéneos) en un vector

```
x <- c(0.5, 0.6); class(x) # numeric
```

```
## [1] "numeric"
```

```
x <- c(TRUE, FALSE); class(x) # logical
```

```
## [1] "logical"
```

```
x <- c(T, F); class(x) # Logical
```

```
## [1] "logical"
```

```
x <- c("a", "b", "c"); class(x) # character
```

```
## [1] "character"
```

```
x <- c(1L, 7L); class(x) # integer
```

```
## [1] "integer"
```

```
x <- c(1+0i, 2+4i); class(x) # complex
```

```
## [1] "complex"
```

- `paste()` (vector de caracteres): Esta función permite concatenar todos los vectores de caracteres que se le suministran y construir una sola cadena de caracteres. También admite argumentos numéricos, que convierte en caracteres. En su forma predeterminada, en la cadena final, cada argumento original se separa del siguiente por un espacio en blanco.

```
args(paste)
```

```
## function (..., sep = " ", collapse = NULL)  
## NULL
```

```
w <- c(0,2,7)  
paste("Raíz de",w,"es",round(sqrt(w),3))
```

```
## [1] "Raíz de 0 es 0"      "Raíz de 2 es 1.414" "Raíz de 7 es 2.646"
```

Esta función es muy útil para crear nombres de variables. Utilizando el argumento `sep` podemos establecer que la separación sea vacía.

```
labs <- paste(c("X","Y"), 1:10, sep="") # No hay separación entre argumentos  
labs; class(labs)
```

```
## [1] "X1" "Y2" "X3" "Y4" "X5" "Y6" "X7" "Y8" "X9" "Y10"
```

```
## [1] "character"
```

- operador : (vector numérico): crea una secuencia de números enteros consecutivos *from:to*

```
x <- 1:10; x; class(x)    #orden ascendente
```

```
## [1] 1 2 3 4 5 6 7 8 9 10
```

```
## [1] "integer"
```

```
x <- 10:1; x; class(x)    #orden descendente
```

```
## [1] 10 9 8 7 6 5 4 3 2 1
```

```
## [1] "integer"
```

```
x <- 2.5*1:5; x; class(x) #Este operador tiene prioridad en una expresión (en ausencia de paréntesis)
```

```
## [1] 2.5 5.0 7.5 10.0 12.5
```

```
## [1] "numeric"
```

- secuencia seq() (vector numérico): función que crea una secuencia regular

```
args(seq.default)
```

```
## function (from = 1, to = 1, by = ((to - from)/(length.out - 1)),  
##     length.out = NULL, along.with = NULL, ...)  
## NULL
```

¿Cómo pasamos los argumentos de las funciones? Los argumentos que tienen nombre, se asignan de acuerdo al nombre y pueden colocarse en cualquier posición. Los valores predeterminados no necesitan ser especificados.

```
x <- seq(from=1, to=10, by=1); x; class(x)
```

```
## [1] 1 2 3 4 5 6 7 8 9 10
```

```
## [1] "numeric"
```

```
x <- seq(by=2,1,7); x
```

```
## [1] 1 3 5 7
```

Si no especificamos el nombre del argumento, se asignan de acuerdo a su posición.

```
x <- seq(0,10,5); x
```

```
## [1] 0 5 10
```

Completa los nombres parciales que usamos para los argumentos.

```
x <- seq(0,10,len=5); x
```

```
## [1] 0.0 2.5 5.0 7.5 10.0
```

- repetición `rep()` (vector numérico): función que crea una secuencia repetida

```
rep(x, times = 1, length.out = NA, each = 1)
```

```
x <- rep(c(1,2),5) ; x; class(x)
```

```
## [1] 1 2 1 2 1 2 1 2 1 2
```

```
## [1] "numeric"
```

```
x <- rep(1:4, c(2,1,2,1)); x
```

```
## [1] 1 1 2 3 3 4
```

```
x <- rep(1:4, each = 2, len = 10); x
```

```
## [1] 1 1 2 2 3 3 4 4 1 1
```

```
x <- rep(1:4, each = 2, times = 3); x
```

```
## [1] 1 1 2 2 3 3 4 4 1 1 2 2 3 3 4 4 1 1 2 2 3 3 4 4
```

- Crear un vector lógico aplicando una condición lógica a otro vector. Los operadores de comparación, lógicos o booleanos son los habituales.

Operador	Descripción
<	menor a
<=	menor o igual a
>	mayor a
>=	mayor o igual a
==	igual a
!=	distinto de
!x	NO x (negación de x)
x & z	x y z (intersección/conjunción)
x   z	x o z (unión/disyunción)
xor(x,y)	OR exclusivo: devuelve TRUE si y solo si un argumento es verdadero

Veamos un ejemplo:

```
x <- -3:4; y <- x>=0; x; y; class(y)
```

```
## [1] -3 -2 -1  0  1  2  3  4
```

```
## [1] FALSE FALSE FALSE  TRUE  TRUE  TRUE  TRUE  TRUE
```

```
## [1] "logical"
```

El vector `y` es un vector de la misma longitud de `x` y cuyos valores serán, respectivamente, `T` o `F` de acuerdo a que los elementos de `x` cumplan o no la condición indicada: ser mayores o iguales a 0.

- `vector()` : crea un vector vacío

```
x <- vector("numeric", length = 10); x; class(x)
```

```
## [1] 0 0 0 0 0 0 0 0 0 0
```

```
## [1] "numeric"
```

## Coerción

En el caso de concatenar objetos de diferentes tipos, **R** reasigna los elementos a la clase apropiada (i.e. la opción que no se presta a ambigüedades) para que así todos los elementos del vector sean de la misma clase.

```
y <- c(1.7, "a"); y; class(y)  # EL resultado de numeric y character es character
```

```
## [1] "1.7" "a"
```

```
## [1] "character"
```

```
y <- c(TRUE, 2); y; class(y)  # EL resultado de logical y numeric es numeric
```

```
## [1] 1 2
```

```
## [1] "numeric"
```

```
y <- c("a", TRUE); y; class(y)  # EL resultado de character y logical es character
```

```
## [1] "a"      "TRUE"
```

```
## [1] "character"
```

En **R** es posible “forzar” (*coerce*) la conversión de una clase de datos en otra, mediante comandos de la forma `as.*`. Por ejemplo:

```
x <- 0:6; class(x)
```

```
## [1] "integer"
```

```
as.numeric(x)
```

```
## [1] 0 1 2 3 4 5 6
```

```
as.logical(x)
```

```
## [1] FALSE TRUE TRUE TRUE TRUE TRUE TRUE
```

```
as.character(x)
```

```
## [1] "0" "1" "2" "3" "4" "5" "6"
```

```
as.complex(x)
```

```
## [1] 0+0i 1+0i 2+0i 3+0i 4+0i 5+0i 6+0i
```

Si proponemos una coerción sin sentido, **R** nos dará como resultado `NA` .

```
x <- c("a", "b", "c")
as.numeric(x)
```

```
## Warning: NAs introduced by coercion
```

```
## [1] NA NA NA
```

```
as.logical(x)
```

```
## [1] NA NA NA
```

## Algunas funciones útiles con vectores

- `length()` : longitud (número de elementos) de un vector
- `names()` : Asigna nombres a los valores de un vector

```
constants <- c(3.1416, 2.7183, 1.4142, 1.6180)
constants
```

```
## [1] 3.14 2.72 1.41 1.62
```

```
names(constants) <- c("pi", "euler", "sqrt2", "golden")
constants
```

```
##      pi  euler sqrt2 golden
##  3.14  2.72  1.41  1.62
```

```
length(constants)
```

```
## [1] 4
```

- `is.vector()` : Devuelve `T` si el objeto es un vector, `F` si no lo es. Como los vectores son atómicos, también pueden ser de utilidad las funciones `is.numeric()` y `is.character()`

```
is.vector(constants); is.numeric(constants); is.character(constants)
```

```
## [1] TRUE
```

```
## [1] TRUE
```

```
## [1] FALSE
```

- Algunas funciones matemáticas (fórmulas que operan vectorialmente): `round()` redondeo al número de decimales requerido; `trunc()` truncación al número entero más próximo; `ceiling()` redondeo al número entero superior; `floor()` redondeo al número entero inferior; `abs()` valor absoluto; `sqrt()` raíz cuadrada; `exp()` función exponencial; `log()` o `ln()` función logarítmica (base e); `log10()` logaritmo en base 10; `scale()` centrado y escalado de una variable.
- Algunas funciones estadísticas (funciones que devuelven estadísticos o resúmenes con información acerca del vector al que se aplican): `sum()` sumatorio de los valores del vector; `prod(x)` productoria; `range()` valores mínimo y máximo del vector; `min()` mínimo; `max()` máximo; `which.max()` índices del valor máximo; `which.min()` índices del valor mínimo; `mean()` media; `sd()` desviación estándar; `var()` varianza; `quantile()` cuartiles de la distribución; `median()` mediana; `IQR()` rango intercuartílico; `mad()` desviación mediana absoluta; `cov()` covarianza; `cor()` correlación; `summary()` resumen del vector, incluye varios valores relevantes (mínimo, máximo, media, mediana, cuartiles).
- `dim()` : dimensión de un objeto en **R**

```
dim(constants)
```

```
## NULL
```

El resultado es simplemente `NULL`, es decir, no tiene una dimensión asociada. Los vectores no tienen dimensión en **R**, más allá de que en álgebra se considera que son estructuras unidimensionales. Esto es, los vectores no son fila o columna, sino que simplemente son una secuencia de valores homogéneos. Sin embargo, cuando se necesita hacer una distinción (por ejemplo, en operaciones aritméticas), **R** normalmente asume que los vectores son columna.

## Indexación de vectores

Usamos los corchetes `[]` para especificar/extraer los elementos de un vector. Para acceder a un dato utilizamos su índice, el cual debe ser un número entero. Así, es posible acceder al valor que ocupa la posición `i` dentro de un vector `x` refiriéndonos a él como `x[i]`.

```
constants[2]
```

```
## euler
## 2.72
```

Si lo que queremos es acceder a un subconjunto de datos podemos utilizar otro vector con los índices correspondientes. Existen distintas formas de indexar vectores:

- Vector de números naturales positivos. El resultado es un vector formado por los elementos del vector referido que corresponden a estos índices y en el orden en que aparecen en el vector de índices.

```
constants[1:2]
```



```
##      pi euler
## 3.14  2.72
```

```
constants[c(1,4,3)]
```

```
##      pi golden  sqrt2
## 3.14  1.62   1.41
```

- Vector de números naturales negativos. En este caso, los índices indican los elementos del vector referido que deben excluirse.

```
constants[c(-1,-2)]
```

```
##      sqrt2 golden
## 1.41   1.62
```

- Vector de caracteres. Esta opción solo puede realizarse si el vector posee el atributo `names()` para identificar sus componentes.

```
constants[c("pi","golden")]
```

```
##      pi golden
## 3.14   1.62
```

- Vector lógico. La expresión lógica será evaluada, componente a componente, como un 0 ( F ) o un 1 ( T ), y se devolverán los datos que corresponden a TRUE en el vector lógico.

```
constants > 2
```

```
##      pi  euler  sqrt2 golden
## TRUE   TRUE  FALSE  FALSE
```

```
as.numeric(constants>2)
```

```
## [1] 1 1 0 0
```

```
constants[constants > 2]
```

```
##      pi euler
## 3.14  2.72
```

## Operaciones con vectores

Ya conocemos las operaciones lógicas. Las operaciones aritméticas son: suma + , resta - , multiplicación \* , división / , potenciación ^ , y división entera ( %/% devuelve el cociente entero de la división y %% el módulo o resto de la división). Por ejemplo, `5/%2` es 2 y `5%%2` es 1 .

En **R**, las operaciones básicas (lógicas o aritméticas) o funciones matemáticas están definidas para operar vectorialmente, componente a componente. En cambio, las funciones estadísticas devuelven por lo general un solo valor para cada vector.

Los vectores lógicos pueden utilizarse en operaciones aritméticas, en cuyo caso se transforman primero en vectores numéricos, de tal modo que F se transforma en 0 y T en 1.

```
a <- 1:4; a
```

```
## [1] 1 2 3 4
```

```
b <- constants>2; b
```

```
##      pi  euler  sqrt2 golden  
##  TRUE   TRUE  FALSE  FALSE
```

```
a*b
```

```
##      pi  euler  sqrt2 golden  
##      1      2      0      0
```

Regla de reciclaje

[COMPLETAR]

## Valores faltantes (*missing values*)

[COMPLETAR]

## Factores

[COMPLETAR]

## Matrices

[COMPLETAR: `as.matrix()`; `t()`]

```
args("matrix")
```

```
## function (data = NA, nrow = 1, ncol = 1, byrow = FALSE, dimnames = NULL)  
## NULL
```

```
# generates 5 x 4 numeric matrix  
mymatrix1 <-matrix(1:20, nrow=5,ncol=4)  
mymatrix1
```

```
##      [,1] [,2] [,3] [,4]  
## [1,]    1    6   11   16  
## [2,]    2    7   12   17  
## [3,]    3    8   13   18  
## [4,]    4    9   14   19  
## [5,]    5   10   15   20
```

```
# another example
cells <- c(1,26,24,68)
rnames <- c("R1", "R2")
cnames <- c("C1", "C2")
mymatrix2 <- matrix(cells, nrow=2, ncol=2, byrow=TRUE,
  dimnames=list(rnames, cnames))
mymatrix2
```

```
##      C1 C2
## R1   1 26
## R2  24 68
```

## Hojas de datos (*Data frames*)

[COMPLETAR]

## Listas

[COMPLETAR]

## Series temporales

La construcción de objetos de la clase `ts()` requiere que los datos de partida estén distribuidos regularmente en la escala temporal utilizada; por ejemplo, que haya siempre un dato por día, o dos datos por mes o un dato cada dos años. En caso de que la serie temporal a tratar esté constituida por valores irregularmente distribuidos en el tiempo deberemos utilizar objetos de la clase `zoo()`.

## Fechas/horas

Podemos convertir los caracteres en fechas mediante el comando `as.Date(x,format)`. Podemos usar los siguientes símbolos para dar formato a las fechas:

Símbolo	Significado	Ejemplo
%d	día (numérico, de 0 a 31)	01-31
%a	día de la semana abreviado a tres letras	Mon
%A	día de la semana (nombre completo)	Monday
%m	mes (numérico de 0 a 12)	00-12
%b	mes (nombre abreviado a tres letras)	Jan
%B	mes (nombre completo)	January
%y	año (con dos dígitos)	07
%Y	año (con cuatro dígitos)	2007

El formato predeterminado de las fechas es “año-mes-día” (**yyyy-mm-dd**).

```
mydates <- as.Date(c("2007-06-22", "2004-02-13"))
class(mydates)
```

```
## [1] "Date"
```

```
mydates
```

```
## [1] "2007-06-22" "2004-02-13"
```

```
strDates <- c("01/05/1965", "08/16/1975")
mydates1 <- as.Date(strDates, "%m/%d/%Y")
class(mydates1)
```

```
## [1] "Date"
```

```
mydates1
```

```
## [1] "1965-01-05" "1975-08-16"
```

## Lectura y escritura de datos

[COMPLETAR: load(); read.csv; read.txt; libreria excel; Foreign and Hsmisc ]

## Exploración de bases de datos

[COMPLETAR: View()]

```
str(Aids2)           # Estructura (resumida) de La base de datos Aids2
```

```
## 'data.frame':    2843 obs. of  7 variables:
## $ state   : Factor w/ 4 levels "NSW","Other",...: 1 1 1 1 1 1 1 1 1 1 ...
## $ sex     : Factor w/ 2 levels "F","M": 2 2 2 2 2 2 2 2 2 2 ...
## $ diag    : int  10905 11029 9551 9577 10015 9971 10746 10042 10464 10439 ...
## $ death   : int  11081 11096 9983 9654 10290 10344 11135 11069 10956 10873 ...
## $ status  : Factor w/ 2 levels "A","D": 2 2 2 2 2 2 2 2 2 2 ...
## $ T.categ: Factor w/ 8 levels "hs","hsid","id",...: 1 1 1 5 1 1 8 1 1 2 ...
## $ age     : int   35 53 42 44 39 36 36 31 26 27 ...
```

```
class(Aids2)         # Clase del objeto (numérico, matriz, data frame, etc.)
```

```
## [1] "data.frame"
```

```
dim(Aids2)           # Dimensión del objeto
```

```
## [1] 2843    7
```

```
names(Aids2)         # Lista de Las variables contenidas en La base de datos Aids2
```

```
## [1] "state"  "sex"    "diag"   "death"  "status" "T.categ" "age"
```

```
head(Aids2)          # Imprime en pantalla Las 6 primeras filas (valor predeterminado)
```

```
##      state sex  diag death status T.categ age
## 1    NSW   M 10905 11081      D      hs  35
## 2    NSW   M 11029 11096      D      hs  53
## 3    NSW   M  9551  9983      D      hs  42
## 4    NSW   M  9577  9654      D    haem  44
## 5    NSW   M 10015 10290      D      hs  39
## 6    NSW   M  9971 10344      D      hs  36
```

```
head(Aids2,n=10)      # Imprime las primeras 10 filas
```

```
##      state sex  diag death status T.categ age
## 1    NSW   M 10905 11081      D      hs  35
## 2    NSW   M 11029 11096      D      hs  53
## 3    NSW   M  9551  9983      D      hs  42
## 4    NSW   M  9577  9654      D    haem  44
## 5    NSW   M 10015 10290      D      hs  39
## 6    NSW   M  9971 10344      D      hs  36
## 7    NSW   M 10746 11135      D    other  36
## 8    NSW   M 10042 11069      D      hs  31
## 9    NSW   M 10464 10956      D      hs  26
## 10   NSW   M 10439 10873      D    hsid  27
```

```
tail(Aids2)           # Imprime las 6 últimas filas (valor predeterminado)
```

```
##      state sex  diag death status T.categ age
## 2838 Other   M 11359 11504      A      hs  27
## 2839 Other   M 11475 11504      A     het  46
## 2840 Other   F 11420 11504      A     het  34
## 2841 Other   M 11496 11504      A    haem  49
## 2842 Other   M 11460 11504      A      hs  55
## 2843 Other   M 11448 11504      A      hs  37
```

```
tail(Aids2,n=10)      # Imprime las 10 últimas filas
```

```
##      state sex  diag death status T.categ age
## 2834 Other   M 11311 11311      D      hs  39
## 2835 Other   M 11297 11504      A      hs  34
## 2836 Other   M 11385 11504      A      hs  37
## 2837 Other   M 11339 11444      D      hs  39
## 2838 Other   M 11359 11504      A      hs  27
## 2839 Other   M 11475 11504      A     het  46
## 2840 Other   F 11420 11504      A     het  34
## 2841 Other   M 11496 11504      A    haem  49
## 2842 Other   M 11460 11504      A      hs  55
## 2843 Other   M 11448 11504      A      hs  37
```

```
levels(Aids2$status) # Lista de los niveles de la variable (factor) status contenida en Aids2
```

```
## [1] "A" "D"
```

```
summary(Aids2)        # Estadísticos descriptivos básicos
```

```
##      state      sex      diag      death      status
## NSW :1780 F: 89 Min. : 8302 Min. : 8469 A:1082
## Other: 249 M:2754 1st Qu.:10163 1st Qu.:10672 D:1761
## QLD : 226 Median :10665 Median :11235
## VIC : 588 Mean :10584 Mean :10990
##      3rd Qu.:11103 3rd Qu.:11504
##      Max. :11503 Max. :11504
##
##      T.categ      age
## hs :2465 Min. : 0.0
## blood : 94 1st Qu.:30.0
## hsid : 72 Median :37.0
## other : 70 Mean :37.4
## id : 48 3rd Qu.:43.0
## haem : 46 Max. :82.0
## (Other): 48
```

Si lo que queremos es visualizar la base de datos al completo en pantalla simplemente escribimos su nombre.

```
Aids2
```

## Manipulación de datos

El paquete `dplyr` es muy útil para la manipulación de datos. Está enfocado a trabajar con *data frames*, principal estructura de datos en **R**.

```
install.packages("dplyr")
library(dplyr)
```

Las funciones principales que provee el paquete `dplyr` son:

- `select()` : devuelve solo las columnas indicadas.
- `filter()` : permite filtrar filas según una expresión lógica.
- `arrange()` : ordena las filas en función de los valores de una o más columnas.
- `rename()` : permite cambiar el nombre de una columna.
- `mutate()` : permite agregar una nueva columna o transformar una existente.
- `summarise()` : permite realizar resúmenes estadísticos de variables.

En el icono *Help* que aparece al final de la barra de menú principal podemos encontrar un chuletero sobre el uso de esta librería pulsando sobre *Cheatsheets*.

## Representación gráfica



[COMPLETAR: Paquete ggplot2]

# Enlaces a bases de datos de interés

De libre acceso:

- Banco de España
  - <https://www.bde.es/bde/es/areas/estadis/> (<https://www.bde.es/bde/es/areas/estadis/>)
  - <https://www.bde.es/investigador/en/> (<https://www.bde.es/investigador/en/>)
- Bases de datos de series de coyuntura económica (BDSICE) - Ministerio de Economía, Industria y Competitividad
  - <http://serviciosede.mineco.gob.es/Indeco/BDSICE/indiceBDSICE.aspx>  
(<http://serviciosede.mineco.gob.es/Indeco/BDSICE/indiceBDSICE.aspx>)
- Instituto Nacional de Estadística (INEbase)
  - <http://www.ine.es/dyngs/INEbase/listaoperaciones.htm>  
(<http://www.ine.es/dyngs/INEbase/listaoperaciones.htm>)
  - <http://www.ine.es/prodyser/microdatos.htm> (<http://www.ine.es/prodyser/microdatos.htm>)
- Eurostat
  - <http://ec.europa.eu/eurostat/data/database> (<http://ec.europa.eu/eurostat/data/database>)
- Banco Mundial
  - <https://data.worldbank.org/indicator> (<https://data.worldbank.org/indicator>)
- Fondo Monetario Internacional
  - <http://www.imf.org/en/Data> (<http://www.imf.org/en/Data>)

Acceso a través de la Biblioteca de la UCM:

- <http://biblioteca.ucm.es/cee/bases-de-datos-de-economia> (<http://biblioteca.ucm.es/cee/bases-de-datos-de-economia>)

## Referencias

- Ballesteros, J. (15-febrero-2018). Introducción a R. Recuperado de [https://rstudio-pubs-static.s3.amazonaws.com/117035\\_02858b3d08a2495cb3c9e3adf2f35c77.html](https://rstudio-pubs-static.s3.amazonaws.com/117035_02858b3d08a2495cb3c9e3adf2f35c77.html) ([https://rstudio-pubs-static.s3.amazonaws.com/117035\\_02858b3d08a2495cb3c9e3adf2f35c77.html](https://rstudio-pubs-static.s3.amazonaws.com/117035_02858b3d08a2495cb3c9e3adf2f35c77.html))
- Quick-R (15-febrero-2018). Recuperado de <https://www.statmethods.net/index.html> (<https://www.statmethods.net/index.html>)
- Santana, A. & Hernández, C.N. (15-febrero-2018). R4ULPGC: Introducción a R. Recuperado de <http://www.dma.ulpgc.es/profesores/personal/stat/cursoR4ULPGC/index.html> (<http://www.dma.ulpgc.es/profesores/personal/stat/cursoR4ULPGC/index.html>)



Proyecto de innovación  
“Innova Docencia” de la UCM  
Convocatoria 2018-2019  
Proyecto número 275

**R-adaptación de la asignatura de métodos econométricos en economía y finanzas del  
Grado de Estadística Aplicada**

Facultad de Estudios Estadísticos  
Departamento de Economía Aplicada, Pública y Política

**Título**

***Práctica 1: Lectura de datos desde la web del INE: La identificación del  
ciclo económico y la ley de Okun***

**Correspondiente al**

**Capítulo 1 INTRODUCCIÓN**

- Naturaleza y contenido de la Econometría
- La modelización económica
- Planteamiento del curso

***Práctica 1: Lectura de datos desde la web del INE: La identificación del ciclo  
económico y la ley de Okun***



# P1. - Ley de Okun: Lectura de datos desde la web del INE

## Asignatura Métodos Econométricos en Economía y Finanzas

Profesores  
Loren Escot, Alicia Pérez-Alonso y Julio E. Sandubete

Dpto Economía Aplicada, Pública y Política  
Facultad Estudios Estadísticos - Universidad Complutense de Madrid

Curso académico 2018/2019

### Abstract

En esta práctica se muestra cómo utilizar la librería `pxR` para descargar datos en formato `pc-Axis` directamente desde la página del INE.

Se mostrará cómo pueden desestacionalizarse series con R, y se mostrará a su vez como enlazar series procedentes de distintas bases utilizando las tasas de crecimiento para generar series temporales largas. Los datos del PIB y de Ocupados generados en esta práctica se guardarán en un archivo `okundat.Rdata` que se utilizarán posteriormente en otras prácticas

- 1 Librerías necesarias
- 2 Datos de Producción Índice de Volumen (Base 2010=100)
- 3 Desestacionalización de series en R
- 4 Enlazando series mediante series de Crecimiento
  - 4.1 PIB base 2008
  - 4.2 PIB base 2000
  - 4.3 PIB base 1995
  - 4.4 PIB base 1986
  - 4.5 Enlaze de la serie PIB por tasas de crecimiento
- 5 Descarga de los datos de Ocupados (Contabilidad Nacional \_Trimestral)
  - 5.1 Datos de Empleo (Base 2010=100)
  - 5.2 Datos de Empleo (Base 1995=100)
  - 5.3 Reconstruyo la serie de ocupados por Tasas de crecimiento
- 6 Fusiono las dos series largas en un dataframe

```
# Lectura de datos desde el INE
```

## 1 Librerías necesarias

```
library(pxR) # para leer datos desde archivos .px pc-axis del INE pxR::read.px()
```

```
## Loading required package: stringr
```

```
## Loading required package: reshape2
```

```
## Loading required package: RJSONIO
```

```
## Loading required package: plyr
```

```
library(dplyr) # para filtrar y seleccionar variables
```

```
##  
## Attaching package: 'dplyr'
```

```
## The following objects are masked from 'package:plyr':  
##  
##   arrange, count, desc, failwith, id, mutate, rename, summarise,  
##   summarize
```

```
## The following objects are masked from 'package:stats':  
##  
##   filter, lag
```

```
## The following objects are masked from 'package:base':  
##  
##   intersect, setdiff, setequal, union
```

## 2 Datos de Producción Índice de Volumen (Base 2010=100)

```
## Datos de Producción Índice de Volumen (Base 2010=100) #####
```

```
url<-"https://www.ine.es/jaxiT3/files/t/es/px/28603.px?nocab=1"
```

```
#datos<-read.px(url) # es una lista
```

```
datos<-as.data.frame(read.px(url), stringsAsFactor=FALSE) #como un data.frame  
names(datos)
```

```
## [1] "Periodo" "Niveles.y.tasas"  
## [3] "Agregados.macroeconómicos" "Tipo.de.dato"  
## [5] "value"
```

```
table(datos$Niveles.y.tasas)
```

```
##  
##           Dato base Variación trimestral      Variación anual  
##           2744           2744           2744
```

```
table(datos$Agregados.macroeconómicos)
```

```
##
##                               Producto interior bruto a precios de mercado
##                               588
##                               VABpb Agricultura, ganadería, silvicultura y pesca (A, CNAE 2009)
##                               588
##                               VABpb Industria (B-E, CNAE 2009)
##                               588
##                               VABpb Industria. Industria manufacturera (C, CNAE 2009)
##                               588
##                               VABpb Construcción (F, CNAE 2009)
##                               588
##                               VABpb Servicios (G-T, CNAE 2009)
##                               588
##                               VABpb Servicios. Comercio, transporte y hostelería (G-I, CNAE 2009)
##                               588
##                               VABpb Servicios. Información y comunicaciones (J, CNAE 2009)
##                               588
##                               VABpb Servicios. Actividades financieras y de seguros (K, CNAE 2009)
##                               588
##                               VABpb Servicios. Actividades inmobiliarias (L, CNAE 2009)
##                               588
## VABpb Servicios. Actividades profesionales, científicas y técnicas y otras (M-N, CNAE 2009)
##                               588
##                               VABpb Servicios. Administración pública, educación y sanidad (O-Q, CNAE 2009)
##                               588
## VABpb Servicios. Actividades artísticas, recreativas y otros servicios (R-T, CNAE 2009)
##                               588
##                               Impuestos menos subvenciones sobre los productos
##                               588
```

```
table(datos$Tipo.de.dato)
```

```
##
## Datos no ajustados de estacionalidad y calendario
##                               4116
## Datos ajustados de estacionalidad y calendario
##                               4116
```

```
pib<-datos%>%
  dplyr::filter(Niveles.y.tasas == "Dato base")%>%
  dplyr::filter(Agregados.macroeconomicos=="Producto interior bruto a precios de mercado")%>%
  dplyr::filter(Tipo.de.dato=="Datos ajustados de estacionalidad y calendario")%>%
  dplyr::mutate(ANO=as.numeric(substr(Periodo,1,4)))%>%
  dplyr::mutate(TRIM=as.numeric(substr(Periodo,6,7)))%>%
  dplyr::mutate(tiempo=ANO+TRIM*0.25-0.25)%>%
  dplyr::arrange(tiempo)%>%
  dplyr::select(Periodo, ANO, TRIM, tiempo, value)

head(pib)
```

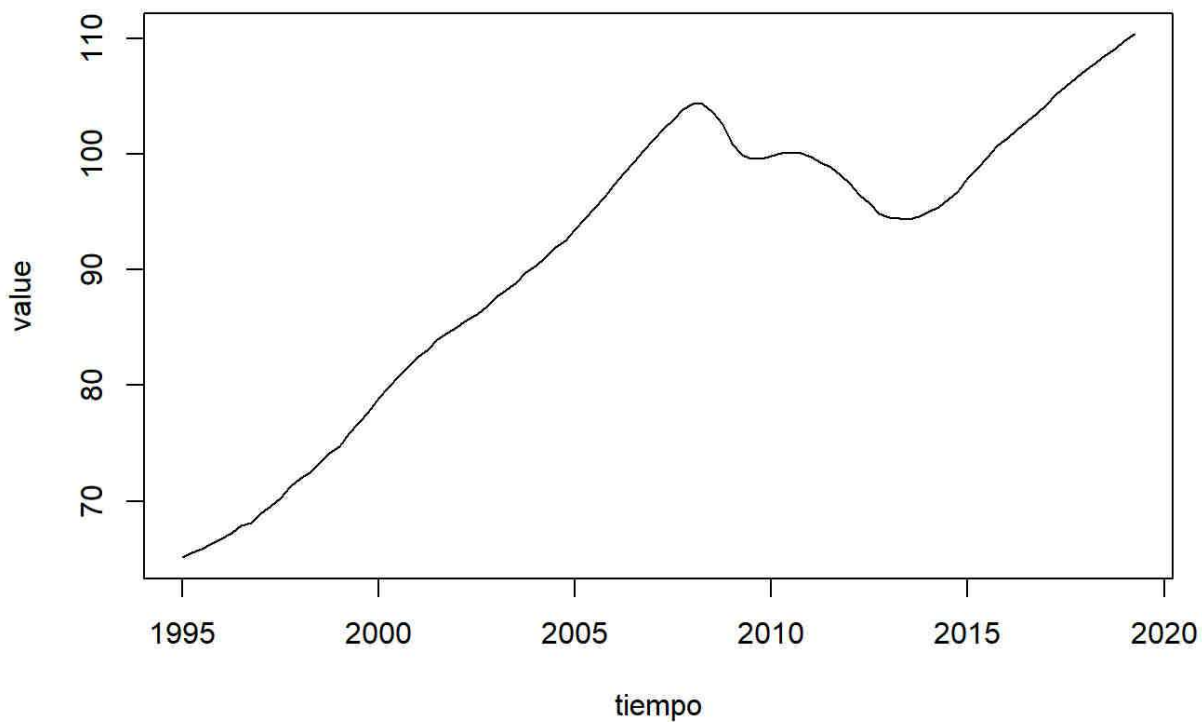
```
## Periodo ANO TRIM tiempo value
## 1 1995T1 1995 1 1995.00 65.1176
## 2 1995T2 1995 2 1995.25 65.5655
## 3 1995T3 1995 3 1995.50 65.8774
## 4 1995T4 1995 4 1995.75 66.3416
## 5 1996T1 1996 1 1996.00 66.7696
## 6 1996T2 1996 2 1996.25 67.1945
```

```
tail(pib)
```

```
##      Periodo ANO TRIM tiempo  value
## 93  2018T1  2018     1 2018.00 107.2630
## 94  2018T2  2018     2 2018.25 107.9003
## 95  2018T3  2018     3 2018.50 108.4781
## 96  2018T4  2018     4 2018.75 109.0791
## 97  2019T1  2019     1 2019.00 109.8059
## 98  2019T2  2019     2 2019.25 110.3294
```

```
tiempo_max<-max(pib$tiempo)
```

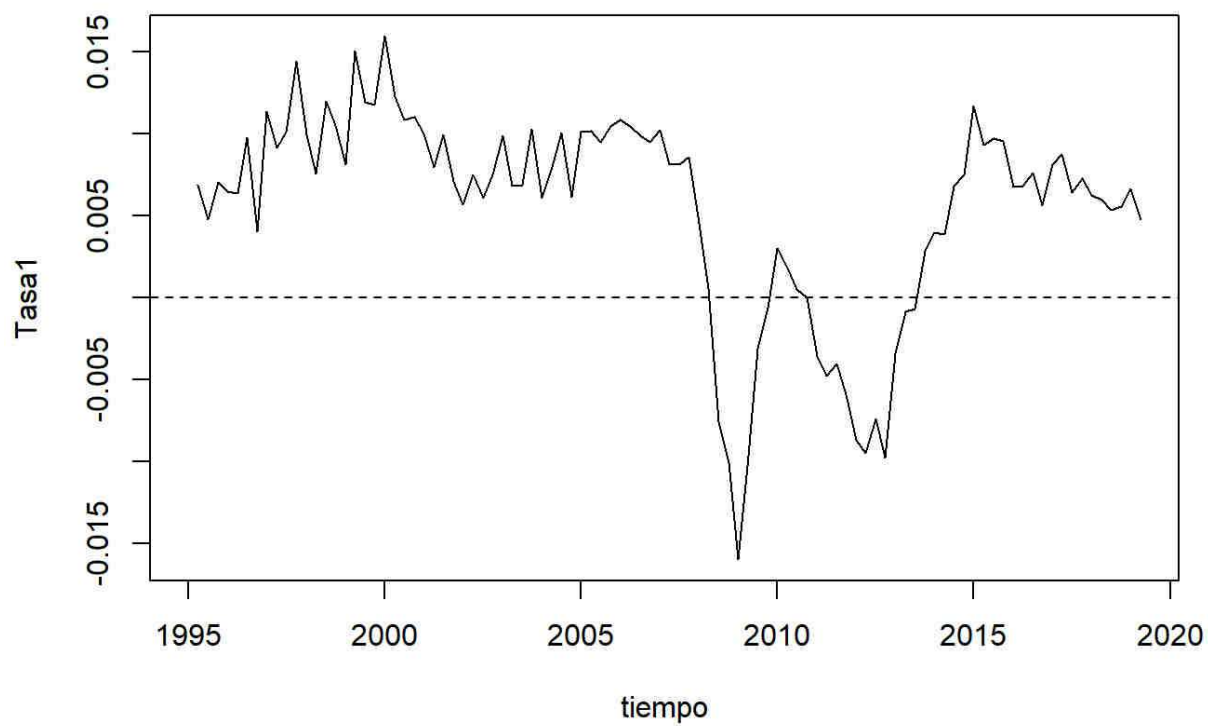
```
plot(value~tiempo,data = pib,type="l")
```



```
# Añado Tasas de Crecimiento
```

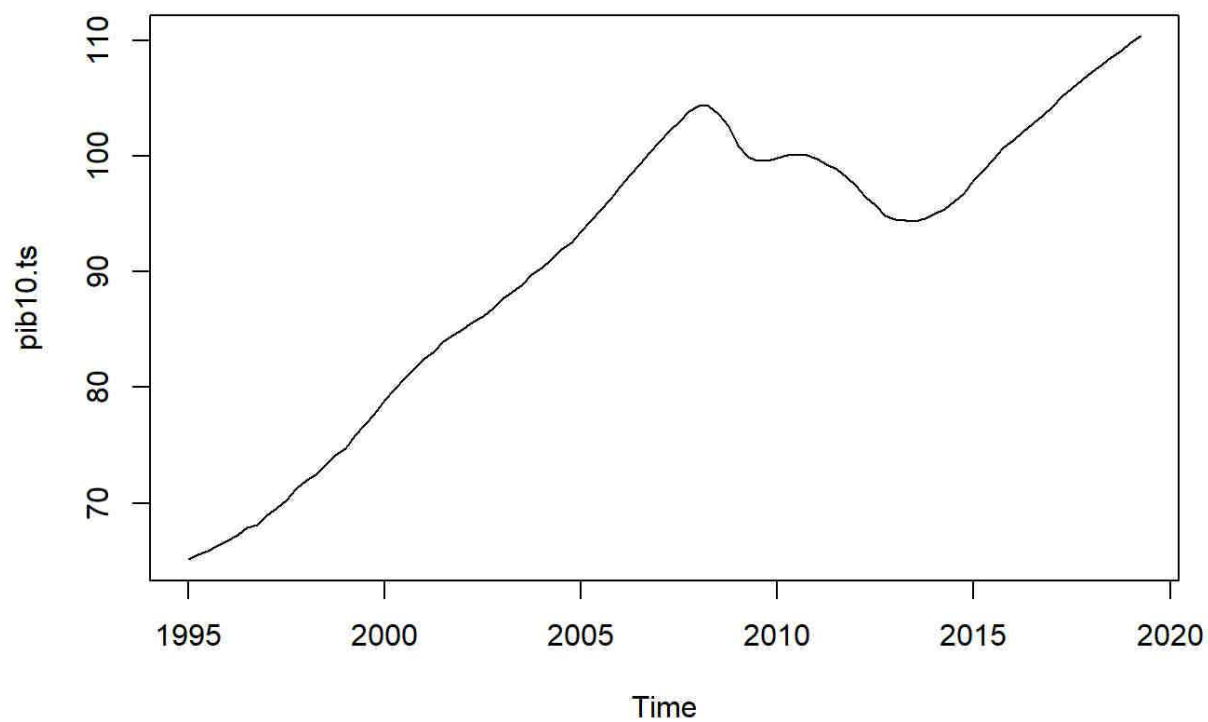
```
pib<-pib%>%
  mutate(Tasa4=value/lag(value,n = 4)-1)%>%
  mutate(Tasa1=value/lag(value,n = 1)-1)
```

```
plot(Tasa1~tiempo, data=pib, type="l")
abline(h=0,lty=2)
```

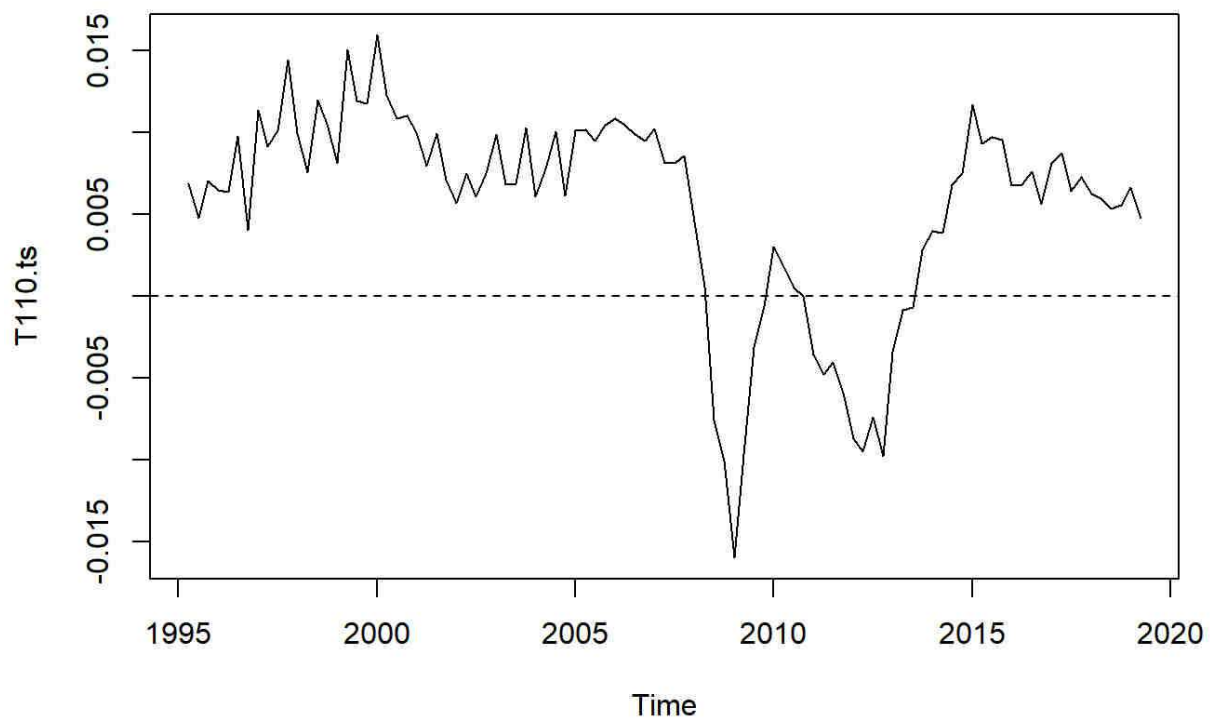


```
#Puedo convertirlo en objeto de serie temporal
pib10.ts<-ts(pib$value, start=c(1995,1), frequency = 4)

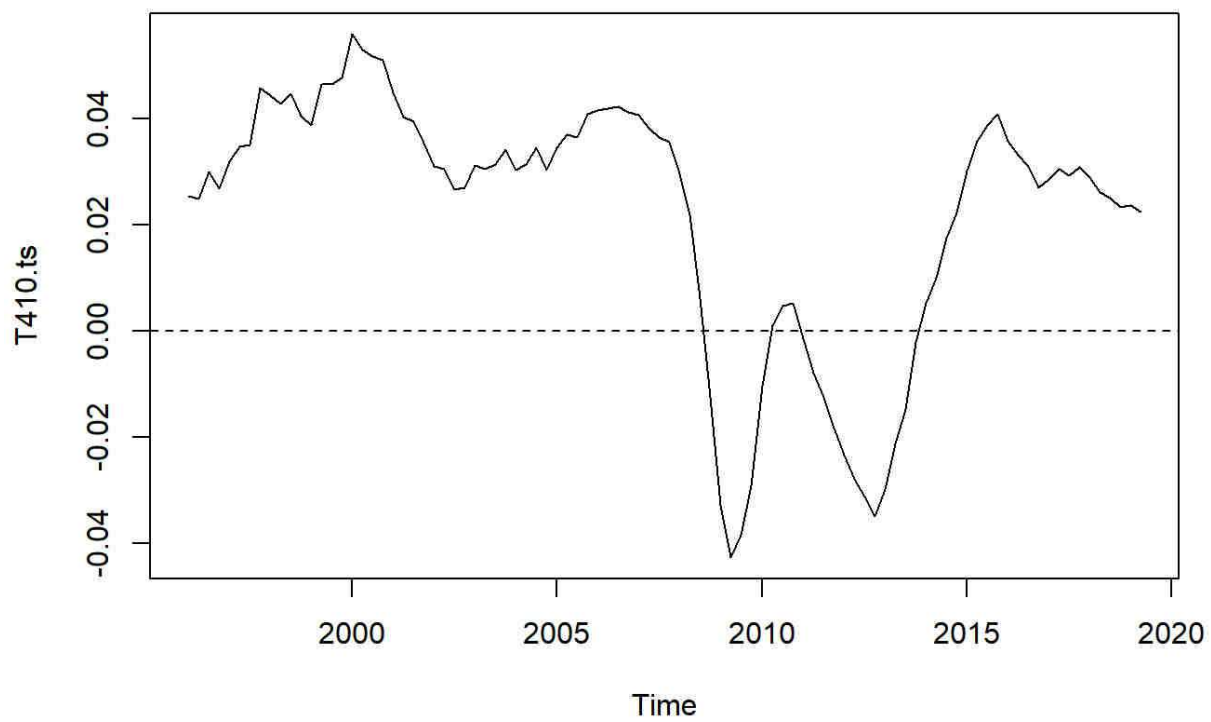
plot(pib10.ts)
```



```
T110.ts<-pib10.ts/stats::lag(pib10.ts, k = -1)-1  
plot(T110.ts)  
abline(h=0,lty=2)
```



```
T410.ts<-pib10.ts/stats::lag(pib10.ts, k = -4)-1  
plot(T410.ts)  
abline(h=0,lty=2)
```



```
ts.union(pib10.ts,T110.ts,T410.ts, dframe = TRUE)
```

##	pib10.ts	T110.ts	T410.ts
## 1	65.1176	NA	NA
## 2	65.5655	6.878325e-03	NA
## 3	65.8774	4.757075e-03	NA
## 4	66.3416	7.046423e-03	NA
## 5	66.7696	6.451457e-03	0.025369485
## 6	67.1945	6.363674e-03	0.024845384
## 7	67.8492	9.743357e-03	0.029931357
## 8	68.1207	4.001521e-03	0.026817261
## 9	68.8962	1.138420e-02	0.031849824
## 10	69.5256	9.135482e-03	0.034691827
## 11	70.2287	1.011282e-02	0.035070421
## 12	71.2431	1.444424e-02	0.045836288
## 13	71.9510	9.936401e-03	0.044339165
## 14	72.4952	7.563481e-03	0.042712325
## 15	73.3644	1.198976e-02	0.044649837
## 16	74.1351	1.050510e-02	0.040593405
## 17	74.7356	8.100077e-03	0.038701338
## 18	75.8624	1.507715e-02	0.046447213
## 19	76.7688	1.194795e-02	0.046403978
## 20	77.6722	1.176780e-02	0.047711543
## 21	78.9107	1.594522e-02	0.055864943
## 22	79.8790	1.227083e-02	0.052945860
## 23	80.7460	1.085392e-02	0.051807505
## 24	81.6371	1.103584e-02	0.051046578
## 25	82.4521	9.983206e-03	0.044878578
## 26	83.1072	7.945219e-03	0.040413626
## 27	83.9345	9.954613e-03	0.039488024
## 28	84.5294	7.087670e-03	0.035428745
## 29	85.0104	5.690328e-03	0.031027712
## 30	85.6495	7.517904e-03	0.030590611
## 31	86.1677	6.050240e-03	0.026606461
## 32	86.8148	7.509775e-03	0.027036747
## 33	87.6715	9.868133e-03	0.031303229
## 34	88.2668	6.790120e-03	0.030558264
## 35	88.8726	6.863283e-03	0.031391113
## 36	89.7854	1.027088e-02	0.034217668
## 37	90.3309	6.075598e-03	0.030333689
## 38	91.0314	7.754821e-03	0.031320950
## 39	91.9487	1.007674e-02	0.034612468
## 40	92.5145	6.153431e-03	0.030395810
## 41	93.4502	1.011409e-02	0.034531927
## 42	94.4028	1.019366e-02	0.037035572
## 43	95.2980	9.482770e-03	0.036425746
## 44	96.2943	1.045457e-02	0.040856298
## 45	97.3373	1.083138e-02	0.041595417
## 46	98.3537	1.044204e-02	0.041851513
## 47	99.3247	9.872531e-03	0.042253772
## 48	100.2681	9.498141e-03	0.041267240
## 49	101.2937	1.022858e-02	0.040646289
## 50	102.1151	8.109093e-03	0.038243604
## 51	102.9436	8.113394e-03	0.036435046
## 52	103.8294	8.604712e-03	0.035517777
## 53	104.3030	4.561328e-03	0.029708659
## 54	104.3612	5.579897e-04	0.021995768
## 55	103.5727	-7.555490e-03	0.006111113
## 56	102.5294	-1.007312e-02	-0.012520538
## 57	100.8896	-1.599346e-02	-0.032725808
## 58	99.9126	-9.683852e-03	-0.042626953
## 59	99.6020	-3.108717e-03	-0.038337322
## 60	99.5395	-6.274974e-04	-0.029161392



```
## 61 99.8398 3.016893e-03 -0.010405433
## 62 100.0214 1.818914e-03 0.001088952
## 63 100.0700 4.858960e-04 0.004698701
## 64 100.0687 -1.299091e-05 0.005316482
## 65 99.7077 -3.607522e-03 -0.001323120
## 66 99.2299 -4.792007e-03 -0.007913307
## 67 98.8268 -4.062284e-03 -0.012423304
## 68 98.2405 -5.932601e-03 -0.018269449
## 69 97.3878 -8.679720e-03 -0.023267009
## 70 96.4614 -9.512485e-03 -0.027899857
## 71 95.7487 -7.388448e-03 -0.031146410
## 72 94.8130 -9.772456e-03 -0.034888870
## 73 94.4877 -3.430964e-03 -0.029778884
## 74 94.4088 -8.350293e-04 -0.021278978
## 75 94.3441 -6.853175e-04 -0.014669651
## 76 94.6135 2.855504e-03 -0.002104142
## 77 94.9875 3.952924e-03 0.005289577
## 78 95.3544 3.862614e-03 0.010016015
## 79 96.0031 6.803042e-03 0.017584565
## 80 96.7233 7.501841e-03 0.022299143
## 81 97.8570 1.172106e-02 0.030209238
## 82 98.7685 9.314612e-03 0.035804326
## 83 99.7277 9.711598e-03 0.038796664
## 84 100.6769 9.517917e-03 0.040875363
## 85 101.3560 6.745341e-03 0.035756257
## 86 102.0443 6.790915e-03 0.033166445
## 87 102.8224 7.625120e-03 0.031031499
## 88 103.4033 5.649547e-03 0.027080691
## 89 104.2456 8.145775e-03 0.028509412
## 90 105.1568 8.740896e-03 0.030501459
## 91 105.8290 6.392359e-03 0.029240710
## 92 106.5982 7.268329e-03 0.030897467
## 93 107.2630 6.236503e-03 0.028945107
## 94 107.9003 5.941471e-03 0.026089611
## 95 108.4781 5.354943e-03 0.025031891
## 96 109.0791 5.540289e-03 0.023273376
## 97 109.8059 6.663055e-03 0.023707150
## 98 110.3294 4.767503e-03 0.022512449
```

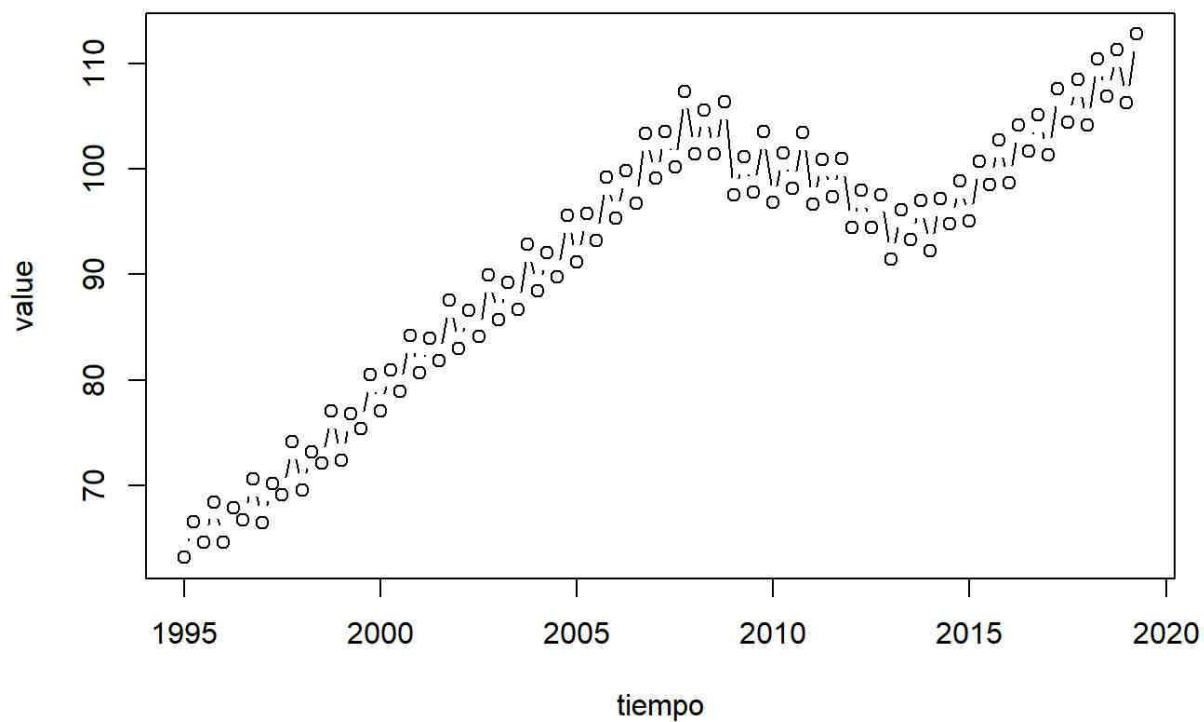
### 3 Desestacionalización de series en R

```
##Desestacionalización de series en R #####
```

```
library (seasonal) #trend(seas(XX)) seas y plot.seas para ajustar las series estacionalmente
```

```
pib_DB<-datos%>%
  filter(Niveles.y.tasas == "Dato base")%>%
  filter(Agregados.macroeconómicos=="Producto interior bruto a precios de mercado")%>%
  filter(Tipo.de.dato=="Datos no ajustados de estacionalidad y calendario")%>%
  mutate(ANO=as.numeric(substr(Periodo,1,4)))%>%
  mutate(TRIM=as.numeric(substr(Periodo,6,7)))%>%
  mutate(tiempo=ANO+TRIM*0.25-0.25)%>%
  arrange(tiempo)%>%
  select(Periodo, ANO, TRIM, tiempo, value)

plot(value~tiempo,data = pib_DB,type="b")
```

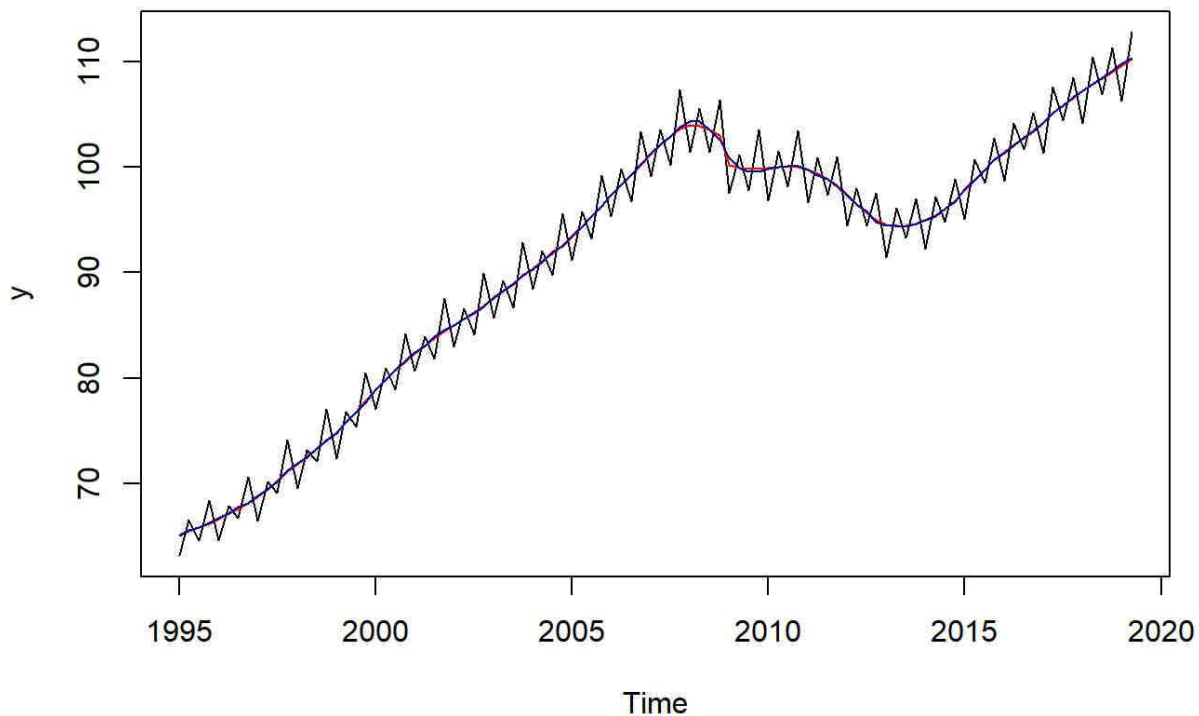


```
# filtro X13
##Componente estacional Seas #librería seasonal::seas, decompose en base
##y_act_trend1<-decompose(na.StructTS(y_interp))$trend #.se limita a hacer una media movil

# hay que convertirla en objeto tipo serie

y<-ts(pib_DB$value, start=c(1995,1), frequency = 4)
y_detrend<-trend(seas(y))

plot(y)
lines(y_detrend, col="red")
lines(pib10.ts,col="darkblue")
```



```
#ydata<-ts.union(y,y_detrend,pib10.ts)
#View(ydata)

##Vamos a unir varias series temporales####
```

## 4 Enlazando series mediante series de Crecimiento

### 4.1 PIB base 2008

```
#PIB base 2008 ----

url<-"http://www.ine.es/jaxiT3/files/t/es/px/2524.px?nocab=1"

datos<-read.px(url) # es una lista

datos<-as.data.frame(read.px(url), stringsAsFactor=FALSE) #como un data.frame
names(datos)
```

```
## [1] "Periodo" "Niveles.y.Tasas"
## [3] "Agregados.macroeconomicos" "Tipo.de.dato"
## [5] "value"
```

```
table(datos$Periodo)
```

```
##
## 2014T2 2014T1 2013T4 2013T3 2013T2 2013T1 2012T4 2012T3 2012T2 2012T1
##      84      84      84      84      84      84      84      84      84      84
## 2011T4 2011T3 2011T2 2011T1 2010T4 2010T3 2010T2 2010T1 2009T4 2009T3
##      84      84      84      84      84      84      84      84      84      84
## 2009T2 2009T1 2008T4 2008T3 2008T2 2008T1 2007T4 2007T3 2007T2 2007T1
##      84      84      84      84      84      84      84      84      84      84
## 2006T4 2006T3 2006T2 2006T1 2005T4 2005T3 2005T2 2005T1 2004T4 2004T3
##      84      84      84      84      84      84      84      84      84      84
## 2004T2 2004T1 2003T4 2003T3 2003T2 2003T1 2002T4 2002T3 2002T2 2002T1
##      84      84      84      84      84      84      84      84      84      84
## 2001T4 2001T3 2001T2 2001T1 2000T4 2000T3 2000T2 2000T1 1999T4 1999T3
##      84      84      84      84      84      84      84      84      84      84
## 1999T2 1999T1 1998T4 1998T3 1998T2 1998T1 1997T4 1997T3 1997T2 1997T1
##      84      84      84      84      84      84      84      84      84      84
## 1996T4 1996T3 1996T2 1996T1 1995T4 1995T3 1995T2 1995T1
##      84      84      84      84      84      84      84      84
```

```
table(datos$Niveles.y.Tasas)
```

```
##
##          Dato base Variación trimestral      Variación anual
##          2184          2184          2184
```

```
table(datos$Agregados.macroeconómicos)
```

```
##
##          Producto interior bruto a precios de mercado
##          468
##          VABpb Agricultura, ganadería, silvicultura y pesca
##          468
##          VABpb Industria
##          468
##          VABpb Industria. Industria manufacturera
##          468
##          VABpb Construcción
##          468
##          VABpb Servicios
##          468
##          VABpb Servicios. Comercio, transporte y hostelería
##          468
##          VABpb Servicios. Información y comunicaciones
##          468
##          VABpb Servicios. Actividades financieras y de seguros
##          468
##          VABpb Servicios. Actividades inmobiliarias
##          468
##          VABpb Servicios. Actividades profesionales
##          468
##          VABpb Servicios. Administración pública, Sanidad y Educación
##          468
##          VABpb Servicios. Actividades artísticas, recreativas y otros servicios
##          468
##          Impuestos menos subvenciones sobre los productos
##          468
```

```
table(datos$Tipo.de.dato)
```

```
##
##                               Datos brutos
##                               3276
## Datos corregidos de efectos estacionales y de calendario
##                               3276
```

```
pib<-datos%>%
  dplyr::filter(Niveles.y.Tasas=="Dato base")%>%
  dplyr::filter(Agregados.macroeconomicos=="Producto interior bruto a precios de mercado")%>%
  dplyr::filter(Tipo.de.dato=="Datos corregidos de efectos estacionales y de calendario")%>%
  dplyr::mutate(ANO=as.numeric(substr(Periodo,1,4)))%>%
  dplyr::mutate(TRIM=as.numeric(substr(Periodo,6,7)))%>%
  dplyr::mutate(tiempo=ANO+TRIM*0.25-0.25)%>%
  dplyr::arrange(tiempo)%>%
  dplyr::select(Periodo, ANO, TRIM, tiempo, value)

head(pib)
```

```
##   Periodo  ANO TRIM  tiempo value
## 1  1995T1 1995    1 1995.00  63.67
## 2  1995T2 1995    2 1995.25  63.98
## 3  1995T3 1995    3 1995.50  64.20
## 4  1995T4 1995    4 1995.75  64.59
## 5  1996T1 1996    1 1996.00  64.94
## 6  1996T2 1996    2 1996.25  65.43
```

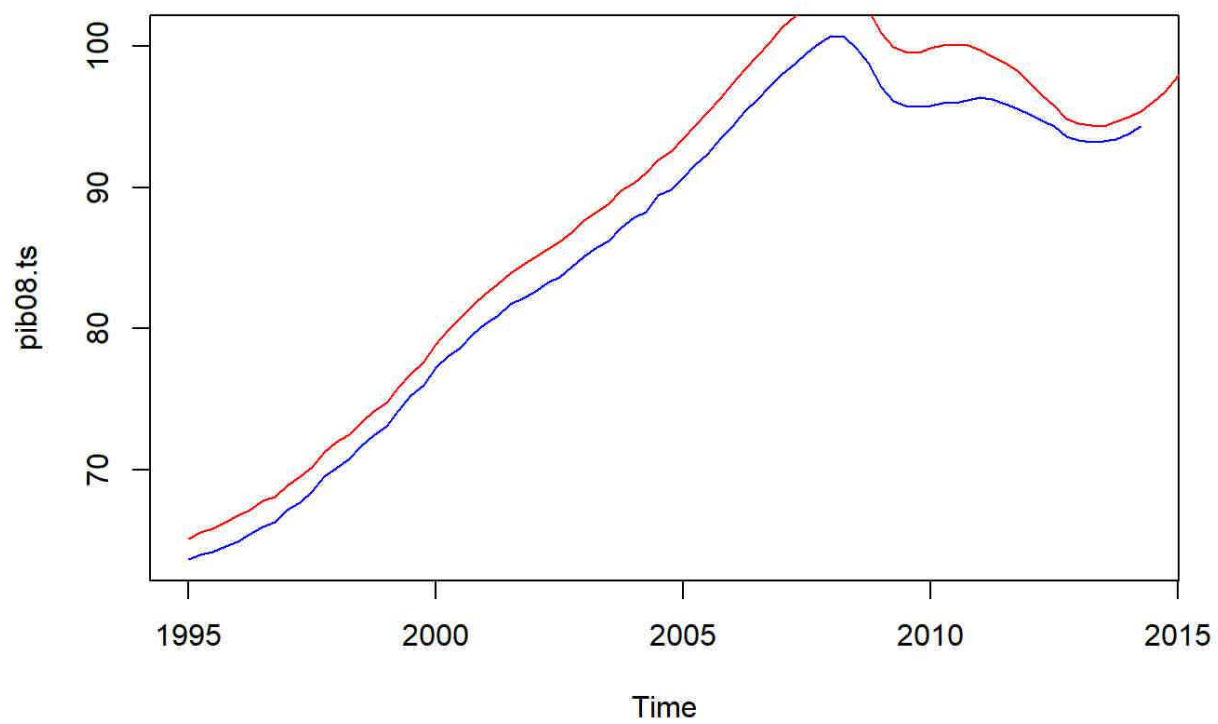
```
tail(pib)
```

```
##   Periodo  ANO TRIM  tiempo value
## 73 2013T1 2013    1 2013.00  93.31
## 74 2013T2 2013    2 2013.25  93.18
## 75 2013T3 2013    3 2013.50  93.26
## 76 2013T4 2013    4 2013.75  93.42
## 77 2014T1 2014    1 2014.00  93.76
## 78 2014T2 2014    2 2014.25  94.29
```

```
#Puedo convertirlo en objeto de serie temporal
pib08.ts<-ts(pib$value, start=c(1995,1), frequency = 4)
T108.ts<-pib08.ts/stats::lag(pib08.ts, k = -1)-1
T408.ts<-pib08.ts/stats::lag(pib08.ts, k = -4)-1

# Dibujo las diferentes series

plot(pib08.ts, col="blue")
lines(pib10.ts, col="red")
```



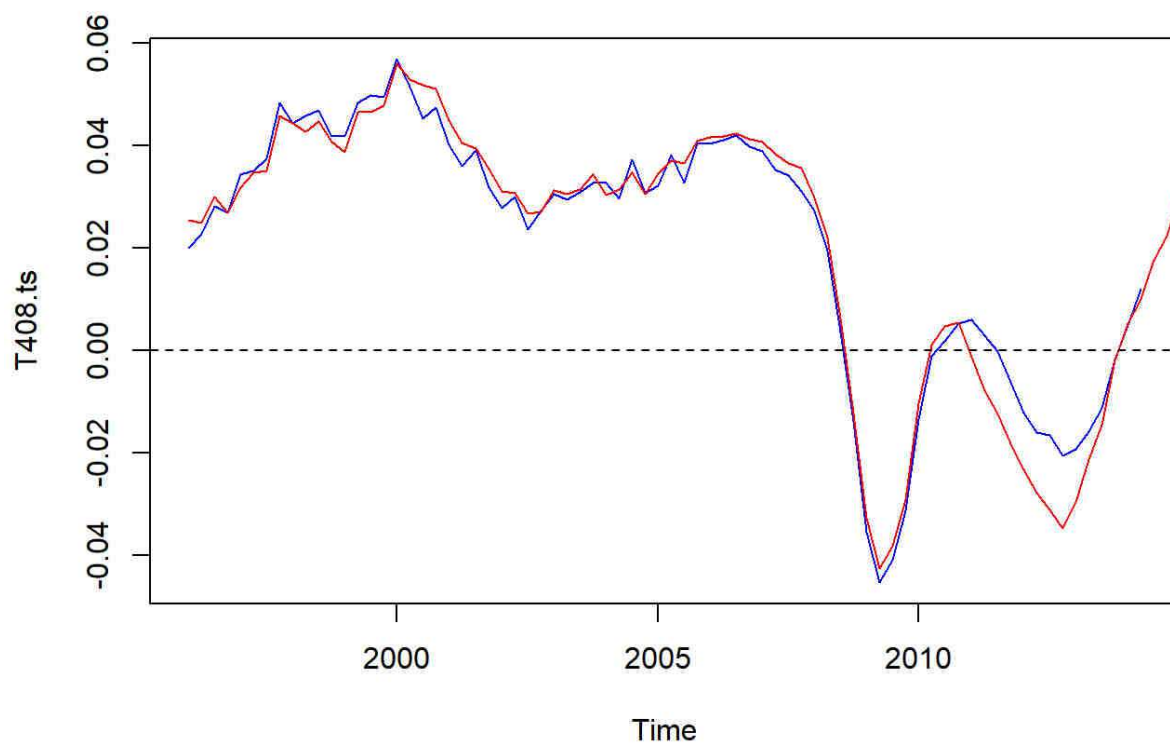
```
ts.union(pib10.ts,pib08.ts)
```

##		pib10.ts	pib08.ts
##	1995 Q1	65.1176	63.67
##	1995 Q2	65.5655	63.98
##	1995 Q3	65.8774	64.20
##	1995 Q4	66.3416	64.59
##	1996 Q1	66.7696	64.94
##	1996 Q2	67.1945	65.43
##	1996 Q3	67.8492	66.01
##	1996 Q4	68.1207	66.33
##	1997 Q1	68.8962	67.17
##	1997 Q2	69.5256	67.72
##	1997 Q3	70.2287	68.48
##	1997 Q4	71.2431	69.53
##	1998 Q1	71.9510	70.15
##	1998 Q2	72.4952	70.82
##	1998 Q3	73.3644	71.69
##	1998 Q4	74.1351	72.43
##	1999 Q1	74.7356	73.08
##	1999 Q2	75.8624	74.24
##	1999 Q3	76.7688	75.26
##	1999 Q4	77.6722	76.01
##	2000 Q1	78.9107	77.23
##	2000 Q2	79.8790	78.05
##	2000 Q3	80.7460	78.66
##	2000 Q4	81.6371	79.61
##	2001 Q1	82.4521	80.33
##	2001 Q2	83.1072	80.86
##	2001 Q3	83.9345	81.73
##	2001 Q4	84.5294	82.14
##	2002 Q1	85.0104	82.56
##	2002 Q2	85.6495	83.28
##	2002 Q3	86.1677	83.65
##	2002 Q4	86.8148	84.37
##	2003 Q1	87.6715	85.08
##	2003 Q2	88.2668	85.73
##	2003 Q3	88.8726	86.23
##	2003 Q4	89.7854	87.13
##	2004 Q1	90.3309	87.86
##	2004 Q2	91.0314	88.27
##	2004 Q3	91.9487	89.44
##	2004 Q4	92.5145	89.81
##	2005 Q1	93.4502	90.69
##	2005 Q2	94.4028	91.63
##	2005 Q3	95.2980	92.36
##	2005 Q4	96.2943	93.45
##	2006 Q1	97.3373	94.34
##	2006 Q2	98.3537	95.39
##	2006 Q3	99.3247	96.24
##	2006 Q4	100.2681	97.16
##	2007 Q1	101.2937	98.01
##	2007 Q2	102.1151	98.75
##	2007 Q3	102.9436	99.52
##	2007 Q4	103.8294	100.18
##	2008 Q1	104.3030	100.68
##	2008 Q2	104.3612	100.66
##	2008 Q3	103.5727	99.87
##	2008 Q4	102.5294	98.79
##	2009 Q1	100.8896	97.11
##	2009 Q2	99.9126	96.09
##	2009 Q3	99.6020	95.78
##	2009 Q4	99.5395	95.69

##	2010	Q1	99.8398	95.79
##	2010	Q2	100.0214	95.97
##	2010	Q3	100.0700	95.95
##	2010	Q4	100.0687	96.18
##	2011	Q1	99.7077	96.35
##	2011	Q2	99.2299	96.24
##	2011	Q3	98.8268	95.93
##	2011	Q4	98.2405	95.57
##	2012	Q1	97.3878	95.16
##	2012	Q2	96.4614	94.69
##	2012	Q3	95.7487	94.33
##	2012	Q4	94.8130	93.60
##	2013	Q1	94.4877	93.31
##	2013	Q2	94.4088	93.18
##	2013	Q3	94.3441	93.26
##	2013	Q4	94.6135	93.42
##	2014	Q1	94.9875	93.76
##	2014	Q2	95.3544	94.29
##	2014	Q3	96.0031	NA
##	2014	Q4	96.7233	NA
##	2015	Q1	97.8570	NA
##	2015	Q2	98.7685	NA
##	2015	Q3	99.7277	NA
##	2015	Q4	100.6769	NA
##	2016	Q1	101.3560	NA
##	2016	Q2	102.0443	NA
##	2016	Q3	102.8224	NA
##	2016	Q4	103.4033	NA
##	2017	Q1	104.2456	NA
##	2017	Q2	105.1568	NA
##	2017	Q3	105.8290	NA
##	2017	Q4	106.5982	NA
##	2018	Q1	107.2630	NA
##	2018	Q2	107.9003	NA
##	2018	Q3	108.4781	NA
##	2018	Q4	109.0791	NA
##	2019	Q1	109.8059	NA
##	2019	Q2	110.3294	NA

```
plot(T408.ts, col="blue")
lines(T410.ts, col="red")
abline(h=0,lty=2)
```





```
ts.union(T408.ts,T410.ts)
```

##		T408.ts	T410.ts
##	1996 Q1	0.0199465997	0.025369485
##	1996 Q2	0.0226633323	0.024845384
##	1996 Q3	0.0281931464	0.029931357
##	1996 Q4	0.0269391547	0.026817261
##	1997 Q1	0.0343393902	0.031849824
##	1997 Q2	0.0349992358	0.034691827
##	1997 Q3	0.0374185729	0.035070421
##	1997 Q4	0.0482436303	0.045836288
##	1998 Q1	0.0443650439	0.044339165
##	1998 Q2	0.0457767277	0.042712325
##	1998 Q3	0.0468750000	0.044649837
##	1998 Q4	0.0417086150	0.040593405
##	1999 Q1	0.0417676408	0.038701338
##	1999 Q2	0.0482914431	0.046447213
##	1999 Q3	0.0497977403	0.046403978
##	1999 Q4	0.0494270330	0.047711543
##	2000 Q1	0.0567870826	0.055864943
##	2000 Q2	0.0513200431	0.052945860
##	2000 Q3	0.0451767207	0.051807505
##	2000 Q4	0.0473621892	0.051046578
##	2001 Q1	0.0401398420	0.044878578
##	2001 Q2	0.0360025625	0.040413626
##	2001 Q3	0.0390287312	0.039488024
##	2001 Q4	0.0317799271	0.035428745
##	2002 Q1	0.0277604880	0.031027712
##	2002 Q2	0.0299282711	0.030590611
##	2002 Q3	0.0234919858	0.026606461
##	2002 Q4	0.0271487704	0.027036747
##	2003 Q1	0.0305232558	0.031303229
##	2003 Q2	0.0294188280	0.030558264
##	2003 Q3	0.0308427974	0.031391113
##	2003 Q4	0.0327130497	0.034217668
##	2004 Q1	0.0326751293	0.030333689
##	2004 Q2	0.0296279016	0.031320950
##	2004 Q3	0.0372260234	0.034612468
##	2004 Q4	0.0307586365	0.030395810
##	2005 Q1	0.0322103346	0.034531927
##	2005 Q2	0.0380650278	0.037035572
##	2005 Q3	0.0326475850	0.036425746
##	2005 Q4	0.0405300078	0.040856298
##	2006 Q1	0.0402469953	0.041595417
##	2006 Q2	0.0410345957	0.041851513
##	2006 Q3	0.0420095279	0.042253772
##	2006 Q4	0.0397003745	0.041267240
##	2007 Q1	0.0389018444	0.040646289
##	2007 Q2	0.0352238180	0.038243604
##	2007 Q3	0.0340814630	0.036435046
##	2007 Q4	0.0310827501	0.035517777
##	2008 Q1	0.0272421182	0.029708659
##	2008 Q2	0.0193417722	0.021995768
##	2008 Q3	0.0035168810	0.006111113
##	2008 Q4	-0.0138750250	-0.012520538
##	2009 Q1	-0.0354588796	-0.032725808
##	2009 Q2	-0.0454003576	-0.042626953
##	2009 Q3	-0.0409532392	-0.038337322
##	2009 Q4	-0.0313796943	-0.029161392
##	2010 Q1	-0.0135928329	-0.010405433
##	2010 Q2	-0.0012488292	0.001088952
##	2010 Q3	0.0017749008	0.004698701
##	2010 Q4	0.0051207023	0.005316482

```
## 2011 Q1 0.0058461217 -0.001323120
## 2011 Q2 0.0028133792 -0.007913307
## 2011 Q3 -0.0002084419 -0.012423304
## 2011 Q4 -0.0063422749 -0.018269449
## 2012 Q1 -0.0123508044 -0.023267009
## 2012 Q2 -0.0161055694 -0.027899857
## 2012 Q3 -0.0166788283 -0.031146410
## 2012 Q4 -0.0206131631 -0.034888870
## 2013 Q1 -0.0194409416 -0.029778884
## 2013 Q2 -0.0159467737 -0.021278978
## 2013 Q3 -0.0113431570 -0.014669651
## 2013 Q4 -0.0019230769 -0.002104142
## 2014 Q1 0.0048226342 0.005289577
## 2014 Q2 0.0119124276 0.010016015
## 2014 Q3 NA 0.017584565
## 2014 Q4 NA 0.022299143
## 2015 Q1 NA 0.030209238
## 2015 Q2 NA 0.035804326
## 2015 Q3 NA 0.038796664
## 2015 Q4 NA 0.040875363
## 2016 Q1 NA 0.035756257
## 2016 Q2 NA 0.033166445
## 2016 Q3 NA 0.031031499
## 2016 Q4 NA 0.027080691
## 2017 Q1 NA 0.028509412
## 2017 Q2 NA 0.030501459
## 2017 Q3 NA 0.029240710
## 2017 Q4 NA 0.030897467
## 2018 Q1 NA 0.028945107
## 2018 Q2 NA 0.026089611
## 2018 Q3 NA 0.025031891
## 2018 Q4 NA 0.023273376
## 2019 Q1 NA 0.023707150
## 2019 Q2 NA 0.022512449
```

## 4.2 PIB base 2000

```
#PIB base 2000 ----
```

```
url<-"http://www.ine.es/jaxiT3/files/t/es/px/2503.px?nocab=1"
```

```
#datos<-read.px(url) # es una lista
```

```
datos<-as.data.frame(read.px(url), stringsAsFactor=FALSE) #como un data.frame
names(datos)
```

```
## [1] "Periodo" "Niveles.y.Tasas"
## [3] "Agregados.macroeconómicos" "Tipo.de.dato"
## [5] "value"
```

```
table(datos$Periodo)
```

```
##
## 2011T2 2011T1 2010T4 2010T3 2010T2 2010T1 2009T4 2009T3 2009T2 2009T1
##      72      72      72      72      72      72      72      72      72      72
## 2008T4 2008T3 2008T2 2008T1 2007T4 2007T3 2007T2 2007T1 2006T4 2006T3
##      72      72      72      72      72      72      72      72      72      72
## 2006T2 2006T1 2005T4 2005T3 2005T2 2005T1 2004T4 2004T3 2004T2 2004T1
##      72      72      72      72      72      72      72      72      72      72
## 2003T4 2003T3 2003T2 2003T1 2002T4 2002T3 2002T2 2002T1 2001T4 2001T3
##      72      72      72      72      72      72      72      72      72      72
## 2001T2 2001T1 2000T4 2000T3 2000T2 2000T1 1999T4 1999T3 1999T2 1999T1
##      72      72      72      72      72      72      72      72      72      72
## 1998T4 1998T3 1998T2 1998T1 1997T4 1997T3 1997T2 1997T1 1996T4 1996T3
##      72      72      72      72      72      72      72      72      72      72
## 1996T2 1996T1 1995T4 1995T3 1995T2 1995T1
##      72      72      72      72      72      72
```

```
table(datos$Niveles.y.Tasas)
```

```
##
##          Dato base Variación trimestral      Variación anual
##          1584          1584          1584
```

```
table(datos$Agregados.macroeconómicos)
```

```
##
##      Producto interior bruto a precios de mercado
##                                396
##      VABpb Agricultura, ganadería y pesca
##                                396
##                                VABpb Energía
##                                396
##                                VABpb Industria
##                                396
##                                VABpb Construcción
##                                396
##                                VABpb Servicios
##                                396
##      VABpb Servicios. Servicios de mercado
##                                396
##      VABpb Servicios. Servicios de no mercado
##                                396
##      Impuestos menos subvenciones sobre los productos
##                                396
##      Impuestos sobre el valor añadido
##                                396
##      Impuestos netos sobre productos importados
##                                396
##      Otros impuestos netos sobre los productos
##                                396
```

```
table(datos$Tipo.de.dato)
```

```
##
##                               Datos brutos
##                               2376
## Datos corregidos de efectos estacionales y de calendario
##                               2376
```

```
pib<-datos%>%
  dplyr::filter(Niveles.y.Tasas == "Dato base")%>%
  dplyr::filter(Agregados.macroeconomicos=="Producto interior bruto a precios de mercado")%>%
  dplyr::filter(Tipo.de.dato=="Datos corregidos de efectos estacionales y de calendario")%>%
  dplyr::mutate(ANO=as.numeric(substr(Periodo,1,4)))%>%
  dplyr::mutate(TRIM=as.numeric(substr(Periodo,6,7)))%>%
  dplyr::mutate(tiempo=ANO+TRIM*0.25-0.25)%>%
  dplyr::arrange(tiempo)%>%
  dplyr::select(Periodo, ANO, TRIM, tiempo, value)

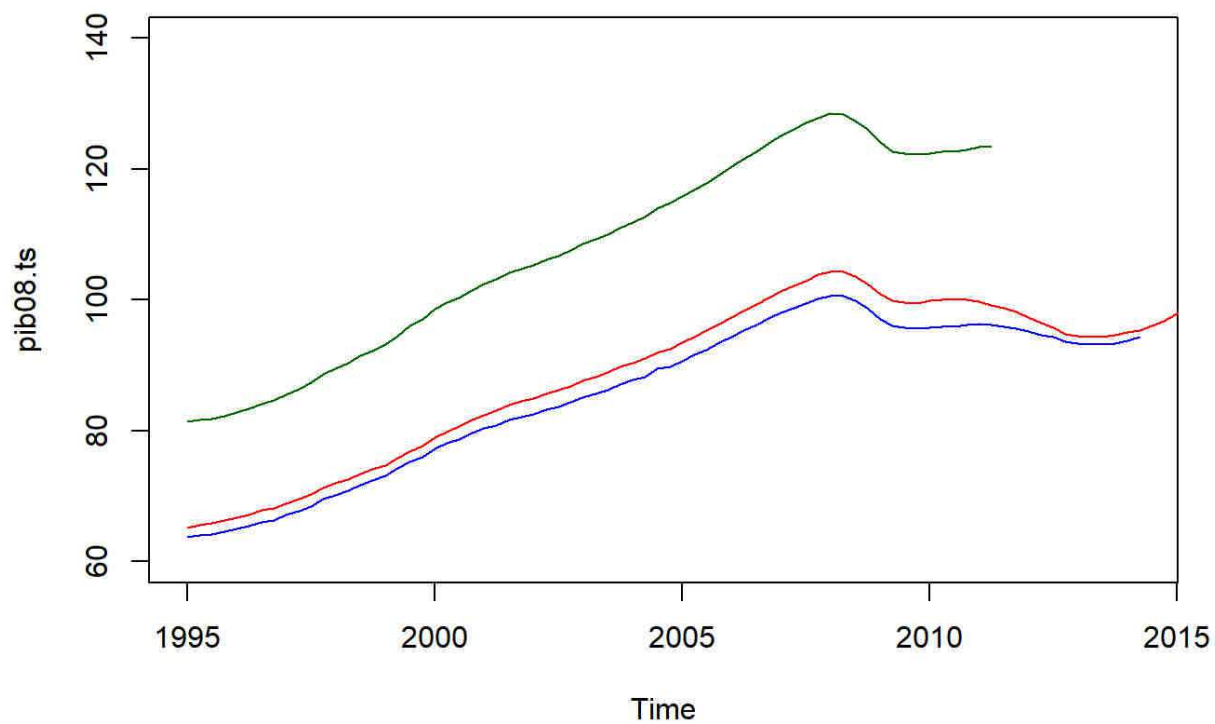
head(pib)
```

```
##   Periodo  ANO TRIM  tiempo value
## 1  1995T1 1995    1 1995.00  81.35
## 2  1995T2 1995    2 1995.25  81.62
## 3  1995T3 1995    3 1995.50  81.85
## 4  1995T4 1995    4 1995.75  82.28
## 5  1996T1 1996    1 1996.00  82.75
## 6  1996T2 1996    2 1996.25  83.44
```

```
#Puedo convertirlo en objeto de serie temporal
pib00.ts<-ts(pib$value, start=c(1995,1), frequency = 4)
T100.ts<-pib00.ts/stats::lag(pib00.ts, k = -1)-1
T400.ts<-pib00.ts/stats::lag(pib00.ts, k = -4)-1

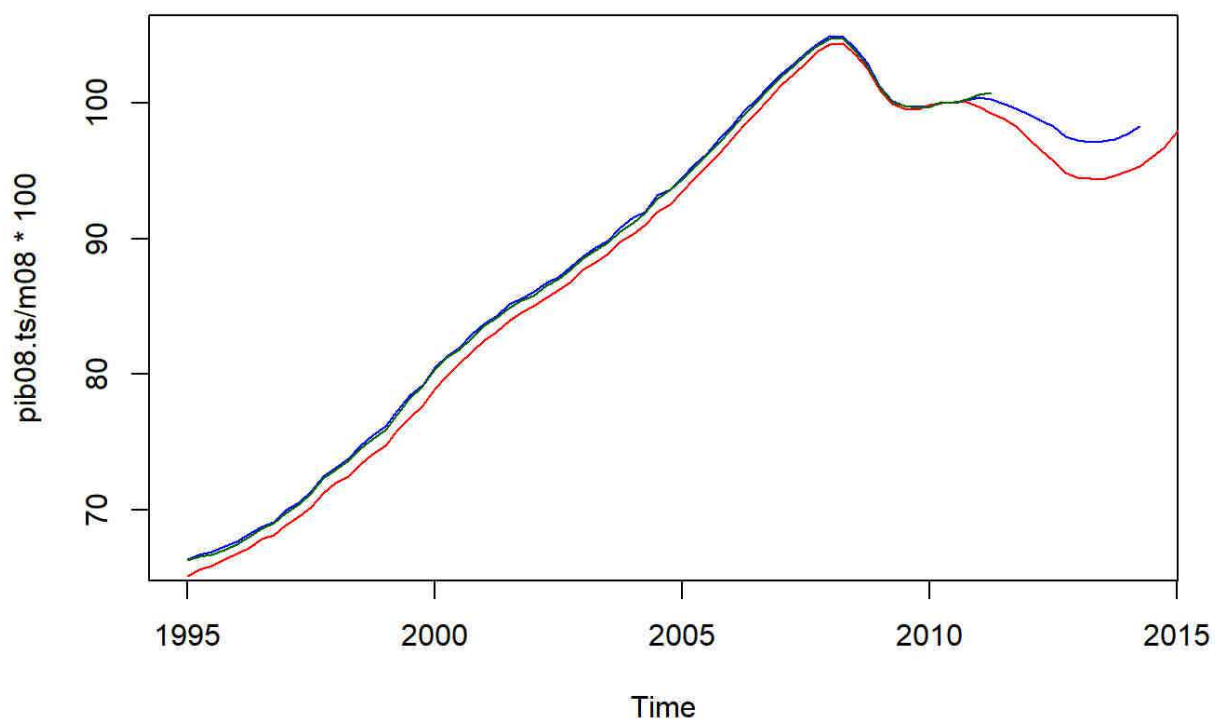
# Dibujo las diferentes series

plot(pib08.ts, col="blue", ylim=c(60,140))
lines(pib10.ts, col="red")
lines(pib00.ts, col="darkgreen")
```



```
m08<-mean(pib08.ts[time(pib08.ts)>=2010 & time(pib08.ts)<2011])
m00<-mean(pib00.ts[time(pib00.ts)>=2010 & time(pib08.ts)<2011])

plot(pib08.ts/m08*100,type="l", col="blue")
lines(pib10.ts,type="l", col="red")
lines(pib00.ts/m00*100,type="l", col="darkgreen")
```



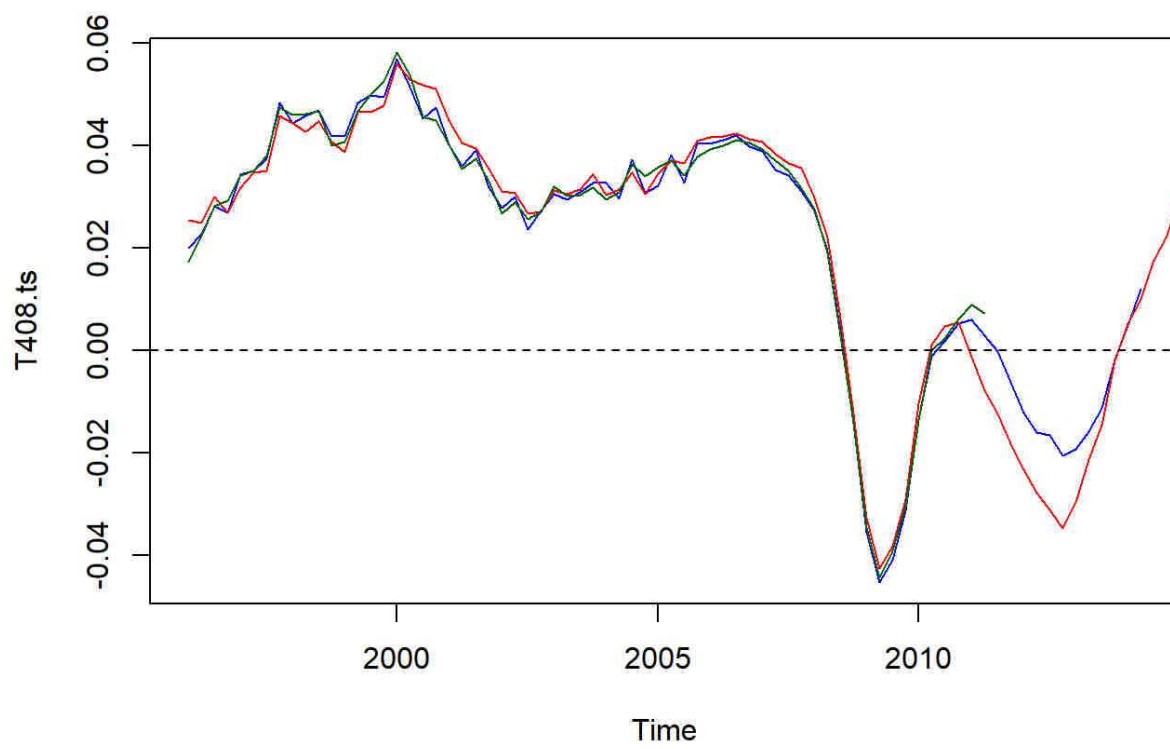
```
ts.union(pib10.ts,pib08.ts,pib00.ts)
```

##		pib10.ts	pib08.ts	pib00.ts
##	1995 Q1	65.1176	63.67	81.35
##	1995 Q2	65.5655	63.98	81.62
##	1995 Q3	65.8774	64.20	81.85
##	1995 Q4	66.3416	64.59	82.28
##	1996 Q1	66.7696	64.94	82.75
##	1996 Q2	67.1945	65.43	83.44
##	1996 Q3	67.8492	66.01	84.15
##	1996 Q4	68.1207	66.33	84.68
##	1997 Q1	68.8962	67.17	85.57
##	1997 Q2	69.5256	67.72	86.36
##	1997 Q3	70.2287	68.48	87.35
##	1997 Q4	71.2431	69.53	88.69
##	1998 Q1	71.9510	70.15	89.50
##	1998 Q2	72.4952	70.82	90.35
##	1998 Q3	73.3644	71.69	91.43
##	1998 Q4	74.1351	72.43	92.24
##	1999 Q1	74.7356	73.08	93.14
##	1999 Q2	75.8624	74.24	94.56
##	1999 Q3	76.7688	75.26	95.99
##	1999 Q4	77.6722	76.01	97.08
##	2000 Q1	78.9107	77.23	98.56
##	2000 Q2	79.8790	78.05	99.65
##	2000 Q3	80.7460	78.66	100.36
##	2000 Q4	81.6371	79.61	101.44
##	2001 Q1	82.4521	80.33	102.51
##	2001 Q2	83.1072	80.86	103.17
##	2001 Q3	83.9345	81.73	104.12
##	2001 Q4	84.5294	82.14	104.79
##	2002 Q1	85.0104	82.56	105.25
##	2002 Q2	85.6495	83.28	106.14
##	2002 Q3	86.1677	83.65	106.79
##	2002 Q4	86.8148	84.37	107.62
##	2003 Q1	87.6715	85.08	108.61
##	2003 Q2	88.2668	85.73	109.33
##	2003 Q3	88.8726	86.23	110.02
##	2003 Q4	89.7854	87.13	111.03
##	2004 Q1	90.3309	87.86	111.81
##	2004 Q2	91.0314	88.27	112.71
##	2004 Q3	91.9487	89.44	114.01
##	2004 Q4	92.5145	89.81	114.80
##	2005 Q1	93.4502	90.69	115.80
##	2005 Q2	94.4028	91.63	116.88
##	2005 Q3	95.2980	92.36	117.91
##	2005 Q4	96.2943	93.45	119.13
##	2006 Q1	97.3373	94.34	120.34
##	2006 Q2	98.3537	95.39	121.55
##	2006 Q3	99.3247	96.24	122.75
##	2006 Q4	100.2681	97.16	123.95
##	2007 Q1	101.2937	98.01	125.06
##	2007 Q2	102.1151	98.75	126.05
##	2007 Q3	102.9436	99.52	127.06
##	2007 Q4	103.8294	100.18	127.87
##	2008 Q1	104.3030	100.68	128.50
##	2008 Q2	104.3612	100.66	128.44
##	2008 Q3	103.5727	99.87	127.42
##	2008 Q4	102.5294	98.79	126.05
##	2009 Q1	100.8896	97.11	124.04
##	2009 Q2	99.9126	96.09	122.73
##	2009 Q3	99.6020	95.78	122.42
##	2009 Q4	99.5395	95.69	122.22



##	2010	Q1	99.8398	95.79	122.33
##	2010	Q2	100.0214	95.97	122.72
##	2010	Q3	100.0700	95.95	122.69
##	2010	Q4	100.0687	96.18	122.96
##	2011	Q1	99.7077	96.35	123.42
##	2011	Q2	99.2299	96.24	123.61
##	2011	Q3	98.8268	95.93	NA
##	2011	Q4	98.2405	95.57	NA
##	2012	Q1	97.3878	95.16	NA
##	2012	Q2	96.4614	94.69	NA
##	2012	Q3	95.7487	94.33	NA
##	2012	Q4	94.8130	93.60	NA
##	2013	Q1	94.4877	93.31	NA
##	2013	Q2	94.4088	93.18	NA
##	2013	Q3	94.3441	93.26	NA
##	2013	Q4	94.6135	93.42	NA
##	2014	Q1	94.9875	93.76	NA
##	2014	Q2	95.3544	94.29	NA
##	2014	Q3	96.0031	NA	NA
##	2014	Q4	96.7233	NA	NA
##	2015	Q1	97.8570	NA	NA
##	2015	Q2	98.7685	NA	NA
##	2015	Q3	99.7277	NA	NA
##	2015	Q4	100.6769	NA	NA
##	2016	Q1	101.3560	NA	NA
##	2016	Q2	102.0443	NA	NA
##	2016	Q3	102.8224	NA	NA
##	2016	Q4	103.4033	NA	NA
##	2017	Q1	104.2456	NA	NA
##	2017	Q2	105.1568	NA	NA
##	2017	Q3	105.8290	NA	NA
##	2017	Q4	106.5982	NA	NA
##	2018	Q1	107.2630	NA	NA
##	2018	Q2	107.9003	NA	NA
##	2018	Q3	108.4781	NA	NA
##	2018	Q4	109.0791	NA	NA
##	2019	Q1	109.8059	NA	NA
##	2019	Q2	110.3294	NA	NA

```
plot(T408.ts, col="blue")
lines(T410.ts, col="red")
lines(T400.ts, col="darkgreen")
abline(h=0,lty=2)
```



```
ts.union(T400.ts,T408.ts,T410.ts)
```

##		T400.ts	T408.ts	T410.ts
##	1996 Q1	1.720959e-02	0.0199465997	0.025369485
##	1996 Q2	2.229846e-02	0.0226633323	0.024845384
##	1996 Q3	2.810018e-02	0.0281931464	0.029931357
##	1996 Q4	2.916869e-02	0.0269391547	0.026817261
##	1997 Q1	3.407855e-02	0.0343393902	0.031849824
##	1997 Q2	3.499521e-02	0.0349992358	0.034691827
##	1997 Q3	3.802733e-02	0.0374185729	0.035070421
##	1997 Q4	4.735475e-02	0.0482436303	0.045836288
##	1998 Q1	4.592731e-02	0.0443650439	0.044339165
##	1998 Q2	4.620195e-02	0.0457767277	0.042712325
##	1998 Q3	4.670864e-02	0.0468750000	0.044649837
##	1998 Q4	4.002706e-02	0.0417086150	0.040593405
##	1999 Q1	4.067039e-02	0.0417676408	0.038701338
##	1999 Q2	4.659657e-02	0.0482914431	0.046447213
##	1999 Q3	4.987422e-02	0.0497977403	0.046403978
##	1999 Q4	5.247181e-02	0.0494270330	0.047711543
##	2000 Q1	5.819197e-02	0.0567870826	0.055864943
##	2000 Q2	5.382826e-02	0.0513200431	0.052945860
##	2000 Q3	4.552558e-02	0.0451767207	0.051807505
##	2000 Q4	4.491141e-02	0.0473621892	0.051046578
##	2001 Q1	4.007711e-02	0.0401398420	0.044878578
##	2001 Q2	3.532363e-02	0.0360025625	0.040413626
##	2001 Q3	3.746513e-02	0.0390287312	0.039488024
##	2001 Q4	3.302445e-02	0.0317799271	0.035428745
##	2002 Q1	2.672910e-02	0.0277604880	0.031027712
##	2002 Q2	2.878744e-02	0.0299282711	0.030590611
##	2002 Q3	2.564349e-02	0.0234919858	0.026606461
##	2002 Q4	2.700639e-02	0.0271487704	0.027036747
##	2003 Q1	3.192399e-02	0.0305232558	0.031303229
##	2003 Q2	3.005464e-02	0.0294188280	0.030558264
##	2003 Q3	3.024628e-02	0.0308427974	0.031391113
##	2003 Q4	3.168556e-02	0.0327130497	0.034217668
##	2004 Q1	2.946322e-02	0.0326751293	0.030333689
##	2004 Q2	3.091558e-02	0.0296279016	0.031320950
##	2004 Q3	3.626613e-02	0.0372260234	0.034612468
##	2004 Q4	3.395479e-02	0.0307586365	0.030395810
##	2005 Q1	3.568554e-02	0.0322103346	0.034531927
##	2005 Q2	3.699760e-02	0.0380650278	0.037035572
##	2005 Q3	3.420753e-02	0.0326475850	0.036425746
##	2005 Q4	3.771777e-02	0.0405300078	0.040856298
##	2006 Q1	3.920553e-02	0.0402469953	0.041595417
##	2006 Q2	3.995551e-02	0.0410345957	0.041851513
##	2006 Q3	4.104826e-02	0.0420095279	0.042253772
##	2006 Q4	4.046000e-02	0.0397003745	0.041267240
##	2007 Q1	3.922220e-02	0.0389018444	0.040646289
##	2007 Q2	3.702180e-02	0.0352238180	0.038243604
##	2007 Q3	3.511202e-02	0.0340814630	0.036435046
##	2007 Q4	3.162566e-02	0.0310827501	0.035517777
##	2008 Q1	2.750680e-02	0.0272421182	0.029708659
##	2008 Q2	1.896073e-02	0.0193417722	0.021995768
##	2008 Q3	2.833307e-03	0.0035168810	0.006111113
##	2008 Q4	-1.423321e-02	-0.0138750250	-0.012520538
##	2009 Q1	-3.470817e-02	-0.0354588796	-0.032725808
##	2009 Q2	-4.445656e-02	-0.0454003576	-0.042626953
##	2009 Q3	-3.924031e-02	-0.0409532392	-0.038337322
##	2009 Q4	-3.038477e-02	-0.0313796943	-0.029161392
##	2010 Q1	-1.378588e-02	-0.0135928329	-0.010405433
##	2010 Q2	-8.147967e-05	-0.0012488292	0.001088952
##	2010 Q3	2.205522e-03	0.0017749008	0.004698701
##	2010 Q4	6.054656e-03	0.0051207023	0.005316482

```
## 2011 Q1 8.910325e-03 0.0058461217 -0.001323120
## 2011 Q2 7.252282e-03 0.0028133792 -0.007913307
## 2011 Q3 NA -0.0002084419 -0.012423304
## 2011 Q4 NA -0.0063422749 -0.018269449
## 2012 Q1 NA -0.0123508044 -0.023267009
## 2012 Q2 NA -0.0161055694 -0.027899857
## 2012 Q3 NA -0.0166788283 -0.031146410
## 2012 Q4 NA -0.0206131631 -0.034888870
## 2013 Q1 NA -0.0194409416 -0.029778884
## 2013 Q2 NA -0.0159467737 -0.021278978
## 2013 Q3 NA -0.0113431570 -0.014669651
## 2013 Q4 NA -0.0019230769 -0.002104142
## 2014 Q1 NA 0.0048226342 0.005289577
## 2014 Q2 NA 0.0119124276 0.010016015
## 2014 Q3 NA NA 0.017584565
## 2014 Q4 NA NA 0.022299143
## 2015 Q1 NA NA 0.030209238
## 2015 Q2 NA NA 0.035804326
## 2015 Q3 NA NA 0.038796664
## 2015 Q4 NA NA 0.040875363
## 2016 Q1 NA NA 0.035756257
## 2016 Q2 NA NA 0.033166445
## 2016 Q3 NA NA 0.031031499
## 2016 Q4 NA NA 0.027080691
## 2017 Q1 NA NA 0.028509412
## 2017 Q2 NA NA 0.030501459
## 2017 Q3 NA NA 0.029240710
## 2017 Q4 NA NA 0.030897467
## 2018 Q1 NA NA 0.028945107
## 2018 Q2 NA NA 0.026089611
## 2018 Q3 NA NA 0.025031891
## 2018 Q4 NA NA 0.023273376
## 2019 Q1 NA NA 0.023707150
## 2019 Q2 NA NA 0.022512449
```

## 4.3 PIB base 1995

```
#PIB base 1995 ----
```

```
url<-"http://www.ine.es/jaxiT3/files/t/es/px/3156.px?nocab=1"
```

```
#datos<-read.px(url) # es una lista
```

```
datos<-as.data.frame(read.px(url), stringsAsFactor=FALSE) #como un data.frame
names(datos)
```

```
## [1] "Periodo" "Niveles.y.Tasas"
## [3] "PIB.a.precios.de.mercado..oferta" "Datos"
## [5] "Precios" "value"
```

```
table(datos$Periodo)
```

```
##
## 2004T4 2004T3 2004T2 2004T1 2003T4 2003T3 2003T2 2003T1 2002T4 2002T3
##      180      180      180      180      180      180      180      180      180      180
## 2002T2 2002T1 2001T4 2001T3 2001T2 2001T1 2000T4 2000T3 2000T2 2000T1
##      180      180      180      180      180      180      180      180      180      180
## 1999T4 1999T3 1999T2 1999T1 1998T4 1998T3 1998T2 1998T1 1997T4 1997T3
##      180      180      180      180      180      180      180      180      180      180
## 1997T2 1997T1 1996T4 1996T3 1996T2 1996T1 1995T4 1995T3 1995T2 1995T1
##      180      180      180      180      180      180      180      180      180      180
## 1994T4 1994T3 1994T2 1994T1 1993T4 1993T3 1993T2 1993T1 1992T4 1992T3
##      180      180      180      180      180      180      180      180      180      180
## 1992T2 1992T1 1991T4 1991T3 1991T2 1991T1 1990T4 1990T3 1990T2 1990T1
##      180      180      180      180      180      180      180      180      180      180
## 1989T4 1989T3 1989T2 1989T1 1988T4 1988T3 1988T2 1988T1 1987T4 1987T3
##      180      180      180      180      180      180      180      180      180      180
## 1987T2 1987T1 1986T4 1986T3 1986T2 1986T1 1985T4 1985T3 1985T2 1985T1
##      180      180      180      180      180      180      180      180      180      180
## 1984T4 1984T3 1984T2 1984T1 1983T4 1983T3 1983T2 1983T1 1982T4 1982T3
##      180      180      180      180      180      180      180      180      180      180
## 1982T2 1982T1 1981T4 1981T3 1981T2 1981T1 1980T4 1980T3 1980T2 1980T1
##      180      180      180      180      180      180      180      180      180      180
```

```
table(datos$Niveles.y.Tasas)
```

```
##
##          Dato base Variación trimestral          Variación anual
##          6000          6000          6000
```

```
table(datos$PIB.a.precios.de.mercado..oferta)
```

```
##
##          Producto interior bruto
##          1800
##          VABpb Agricultura, ganadería y pesca
##          1800
##          VABpb Energía
##          1800
##          VABpb Industria
##          1800
##          VABpb Construcción
##          1800
##          VABpb Servicios. Servicios de mercado
##          1800
##          VABpb Servicios. Servicios de no mercado
##          1800
##          IVA que grava los productos
##          1800
## Impuestos netos sobre productos importados
##          1800
## Otros impuestos netos sobre los productos
##          1800
```

```
table(datos$Datos)
```

```
##
##                               Datos brutos
##                               6000
## Datos corregidos de estacionalidad y calendario
##                               6000
##                               Ciclotendencia
##                               6000
```

```
table(datos$Precios)
```

```
##
## Precios corrientes Precios constantes
##           9000           9000
```

```
pib<-datos%>%
  dplyr::filter(Niveles.y.Tasas == "Dato base")%>%
  dplyr::filter(PIB.a.precios.de.mercado..oferta=="Producto interior bruto")%>%
  dplyr::filter(Datos=="Ciclotendencia")%>%
  dplyr::filter(Precios=="Precios constantes")%>%
  dplyr::mutate(ANO=as.numeric(substr(Periodo,1,4)))%>%
  dplyr::mutate(TRIM=as.numeric(substr(Periodo,6,7)))%>%
  dplyr::mutate(tiempo=ANO+TRIM*0.25-0.25)%>%
  dplyr::arrange(tiempo)%>%
  dplyr::select(Periodo, ANO, TRIM, tiempo, value)

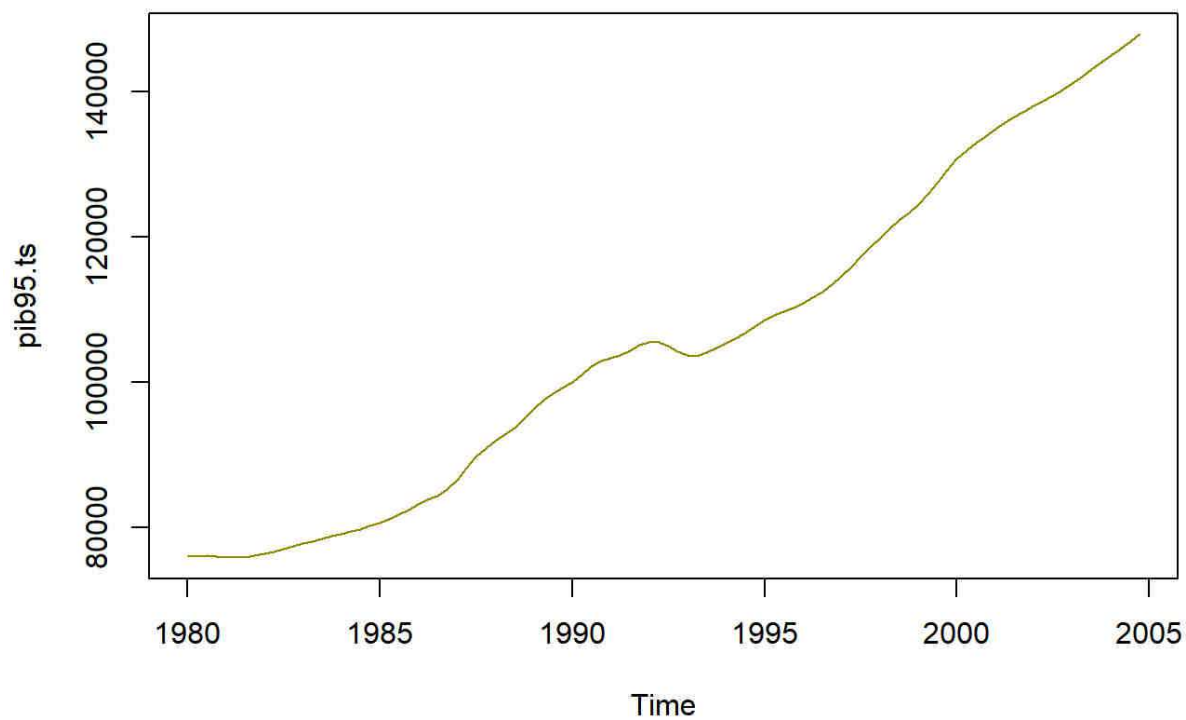
head(pib)
```

```
##   Periodo  ANO TRIM  tiempo value
## 1  1980T1 1980    1 1980.00 76035
## 2  1980T2 1980    2 1980.25 76073
## 3  1980T3 1980    3 1980.50 76090
## 4  1980T4 1980    4 1980.75 76025
## 5  1981T1 1981    1 1981.00 75923
## 6  1981T2 1981    2 1981.25 75865
```

```
#Puedo convertirlo en objeto de serie temporal
pib95.ts<-ts(pib$value, start=c(1980,1), frequency = 4)
T195.ts<-pib95.ts/stats::lag(pib95.ts, k = -1)-1
T495.ts<-pib95.ts/stats::lag(pib95.ts, k = -4)-1

# Dibujo las diferentes series

plot(pib95.ts, col="yellow4")
```

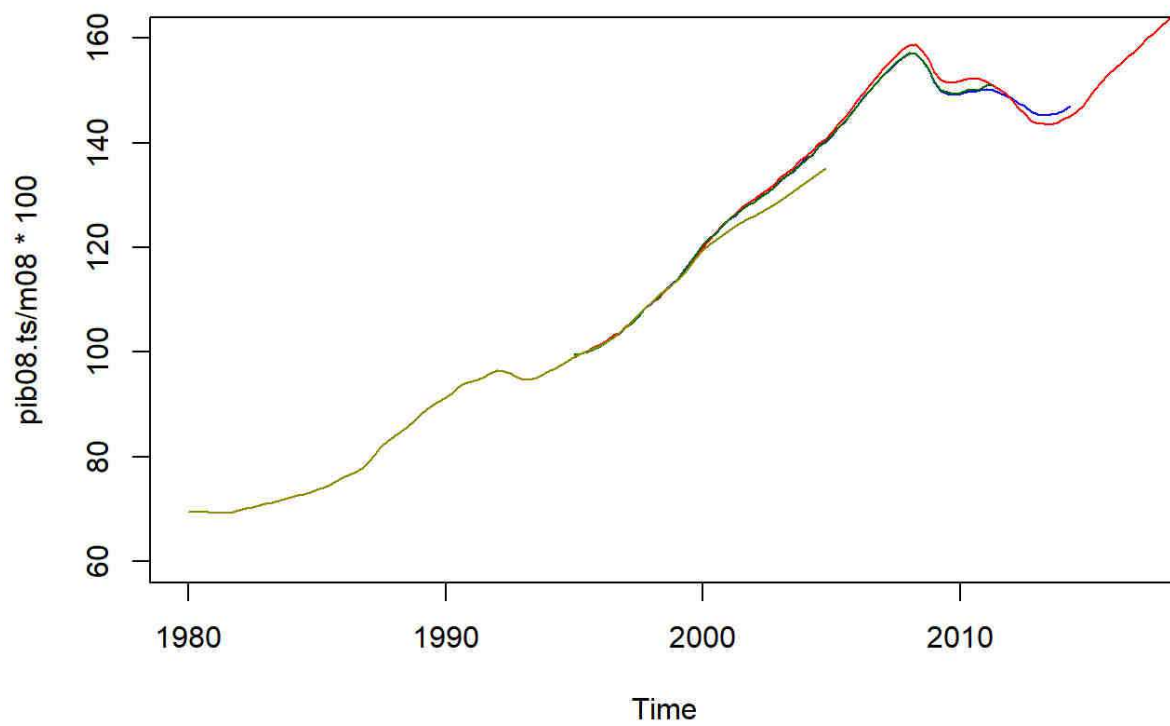


```

m95<-mean(pib95.ts[time(pib95.ts)>=1995 & time(pib95.ts)<1996])
m10<-mean(pib10.ts[time(pib10.ts)>=1995 & time(pib10.ts)<1996])
m08<-mean(pib08.ts[time(pib08.ts)>=1995 & time(pib08.ts)<1996])
m00<-mean(pib00.ts[time(pib00.ts)>=1995 & time(pib00.ts)<1996])

plot(pib08.ts/m08*100,type="l", col="blue", xlim=c(1980,2017), ylim=c(60,160))
lines(pib10.ts/m10*100,type="l", col="red")
lines(pib00.ts/m00*100,type="l", col="darkgreen")
lines(pib95.ts/m95*100,type="l", col="yellow4")

```



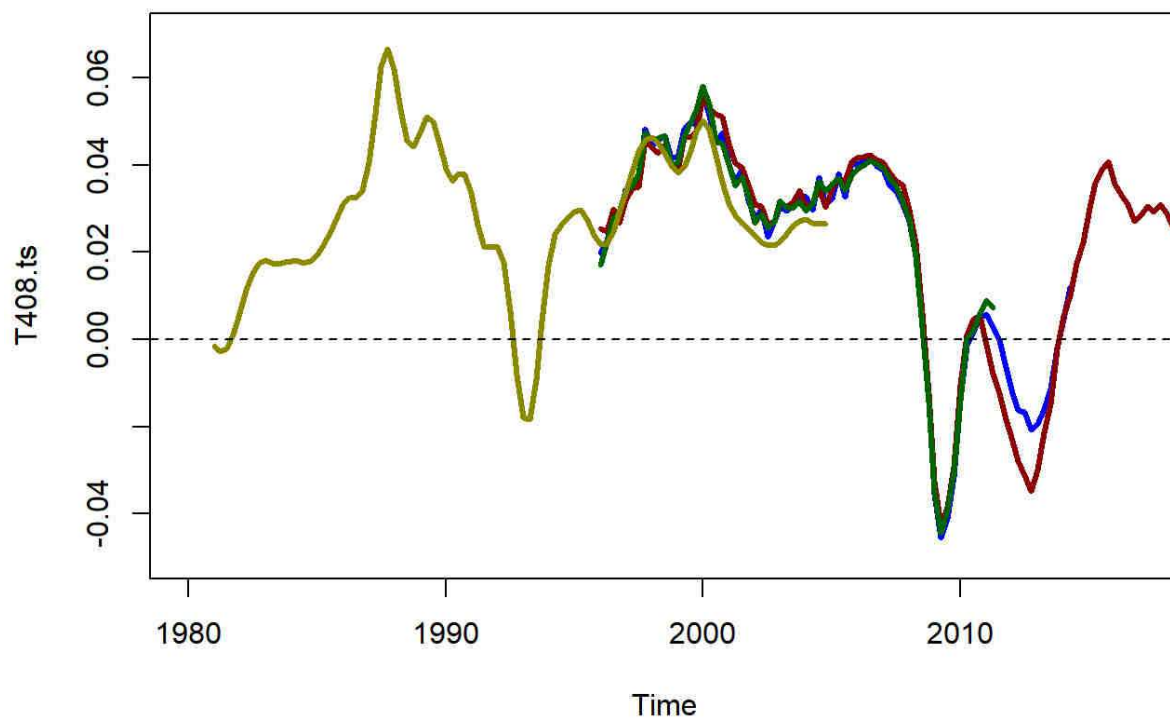
```
ts.union(pib10.ts,pib08.ts,pib00.ts)
```



##		pib10.ts	pib08.ts	pib00.ts
##	1995 Q1	65.1176	63.67	81.35
##	1995 Q2	65.5655	63.98	81.62
##	1995 Q3	65.8774	64.20	81.85
##	1995 Q4	66.3416	64.59	82.28
##	1996 Q1	66.7696	64.94	82.75
##	1996 Q2	67.1945	65.43	83.44
##	1996 Q3	67.8492	66.01	84.15
##	1996 Q4	68.1207	66.33	84.68
##	1997 Q1	68.8962	67.17	85.57
##	1997 Q2	69.5256	67.72	86.36
##	1997 Q3	70.2287	68.48	87.35
##	1997 Q4	71.2431	69.53	88.69
##	1998 Q1	71.9510	70.15	89.50
##	1998 Q2	72.4952	70.82	90.35
##	1998 Q3	73.3644	71.69	91.43
##	1998 Q4	74.1351	72.43	92.24
##	1999 Q1	74.7356	73.08	93.14
##	1999 Q2	75.8624	74.24	94.56
##	1999 Q3	76.7688	75.26	95.99
##	1999 Q4	77.6722	76.01	97.08
##	2000 Q1	78.9107	77.23	98.56
##	2000 Q2	79.8790	78.05	99.65
##	2000 Q3	80.7460	78.66	100.36
##	2000 Q4	81.6371	79.61	101.44
##	2001 Q1	82.4521	80.33	102.51
##	2001 Q2	83.1072	80.86	103.17
##	2001 Q3	83.9345	81.73	104.12
##	2001 Q4	84.5294	82.14	104.79
##	2002 Q1	85.0104	82.56	105.25
##	2002 Q2	85.6495	83.28	106.14
##	2002 Q3	86.1677	83.65	106.79
##	2002 Q4	86.8148	84.37	107.62
##	2003 Q1	87.6715	85.08	108.61
##	2003 Q2	88.2668	85.73	109.33
##	2003 Q3	88.8726	86.23	110.02
##	2003 Q4	89.7854	87.13	111.03
##	2004 Q1	90.3309	87.86	111.81
##	2004 Q2	91.0314	88.27	112.71
##	2004 Q3	91.9487	89.44	114.01
##	2004 Q4	92.5145	89.81	114.80
##	2005 Q1	93.4502	90.69	115.80
##	2005 Q2	94.4028	91.63	116.88
##	2005 Q3	95.2980	92.36	117.91
##	2005 Q4	96.2943	93.45	119.13
##	2006 Q1	97.3373	94.34	120.34
##	2006 Q2	98.3537	95.39	121.55
##	2006 Q3	99.3247	96.24	122.75
##	2006 Q4	100.2681	97.16	123.95
##	2007 Q1	101.2937	98.01	125.06
##	2007 Q2	102.1151	98.75	126.05
##	2007 Q3	102.9436	99.52	127.06
##	2007 Q4	103.8294	100.18	127.87
##	2008 Q1	104.3030	100.68	128.50
##	2008 Q2	104.3612	100.66	128.44
##	2008 Q3	103.5727	99.87	127.42
##	2008 Q4	102.5294	98.79	126.05
##	2009 Q1	100.8896	97.11	124.04
##	2009 Q2	99.9126	96.09	122.73
##	2009 Q3	99.6020	95.78	122.42
##	2009 Q4	99.5395	95.69	122.22

##	2010	Q1	99.8398	95.79	122.33
##	2010	Q2	100.0214	95.97	122.72
##	2010	Q3	100.0700	95.95	122.69
##	2010	Q4	100.0687	96.18	122.96
##	2011	Q1	99.7077	96.35	123.42
##	2011	Q2	99.2299	96.24	123.61
##	2011	Q3	98.8268	95.93	NA
##	2011	Q4	98.2405	95.57	NA
##	2012	Q1	97.3878	95.16	NA
##	2012	Q2	96.4614	94.69	NA
##	2012	Q3	95.7487	94.33	NA
##	2012	Q4	94.8130	93.60	NA
##	2013	Q1	94.4877	93.31	NA
##	2013	Q2	94.4088	93.18	NA
##	2013	Q3	94.3441	93.26	NA
##	2013	Q4	94.6135	93.42	NA
##	2014	Q1	94.9875	93.76	NA
##	2014	Q2	95.3544	94.29	NA
##	2014	Q3	96.0031	NA	NA
##	2014	Q4	96.7233	NA	NA
##	2015	Q1	97.8570	NA	NA
##	2015	Q2	98.7685	NA	NA
##	2015	Q3	99.7277	NA	NA
##	2015	Q4	100.6769	NA	NA
##	2016	Q1	101.3560	NA	NA
##	2016	Q2	102.0443	NA	NA
##	2016	Q3	102.8224	NA	NA
##	2016	Q4	103.4033	NA	NA
##	2017	Q1	104.2456	NA	NA
##	2017	Q2	105.1568	NA	NA
##	2017	Q3	105.8290	NA	NA
##	2017	Q4	106.5982	NA	NA
##	2018	Q1	107.2630	NA	NA
##	2018	Q2	107.9003	NA	NA
##	2018	Q3	108.4781	NA	NA
##	2018	Q4	109.0791	NA	NA
##	2019	Q1	109.8059	NA	NA
##	2019	Q2	110.3294	NA	NA

```
plot(T408.ts, col="blue", xlim=c(1980,2017), ylim=c(-0.05,0.07),lwd=3)
lines(T410.ts, col="darkred",lwd=3)
lines(T400.ts, col="darkgreen",lwd=3)
lines(T495.ts, col="yellow4",lwd=3)
abline(h=0,lty=2)
```



```
ts.union(T495.ts,T400.ts,T408.ts,T410.ts)
```

##		T495.ts	T400.ts	T408.ts	T410.ts
## 1981	Q1	-0.001473006	NA	NA	NA
## 1981	Q2	-0.002734216	NA	NA	NA
## 1981	Q3	-0.002155342	NA	NA	NA
## 1981	Q4	0.001025978	NA	NA	NA
## 1982	Q1	0.005913886	NA	NA	NA
## 1982	Q2	0.011243656	NA	NA	NA
## 1982	Q3	0.015264863	NA	NA	NA
## 1982	Q4	0.017476315	NA	NA	NA
## 1983	Q1	0.018095637	NA	NA	NA
## 1983	Q2	0.017609948	NA	NA	NA
## 1983	Q3	0.017409353	NA	NA	NA
## 1983	Q4	0.017692715	NA	NA	NA
## 1984	Q1	0.018018366	NA	NA	NA
## 1984	Q2	0.017881618	NA	NA	NA
## 1984	Q3	0.017481225	NA	NA	NA
## 1984	Q4	0.017956169	NA	NA	NA
## 1985	Q1	0.019152296	NA	NA	NA
## 1985	Q2	0.021531492	NA	NA	NA
## 1985	Q3	0.024474298	NA	NA	NA
## 1985	Q4	0.027599791	NA	NA	NA
## 1986	Q1	0.030940487	NA	NA	NA
## 1986	Q2	0.032386420	NA	NA	NA
## 1986	Q3	0.032562293	NA	NA	NA
## 1986	Q4	0.034246409	NA	NA	NA
## 1987	Q1	0.040520880	NA	NA	NA
## 1987	Q2	0.051810751	NA	NA	NA
## 1987	Q3	0.062596993	NA	NA	NA
## 1987	Q4	0.066565011	NA	NA	NA
## 1988	Q1	0.061927268	NA	NA	NA
## 1988	Q2	0.052696065	NA	NA	NA
## 1988	Q3	0.045497620	NA	NA	NA
## 1988	Q4	0.044220829	NA	NA	NA
## 1989	Q1	0.047423174	NA	NA	NA
## 1989	Q2	0.050974222	NA	NA	NA
## 1989	Q3	0.049969076	NA	NA	NA
## 1989	Q4	0.044812588	NA	NA	NA
## 1990	Q1	0.039146425	NA	NA	NA
## 1990	Q2	0.036299502	NA	NA	NA
## 1990	Q3	0.038034205	NA	NA	NA
## 1990	Q4	0.037810213	NA	NA	NA
## 1991	Q1	0.033492632	NA	NA	NA
## 1991	Q2	0.026191843	NA	NA	NA
## 1991	Q3	0.021074465	NA	NA	NA
## 1991	Q4	0.021271016	NA	NA	NA
## 1992	Q1	0.021408118	NA	NA	NA
## 1992	Q2	0.017587673	NA	NA	NA
## 1992	Q3	0.006429482	NA	NA	NA
## 1992	Q4	-0.007988816	NA	NA	NA
## 1993	Q1	-0.017815031	NA	NA	NA
## 1993	Q2	-0.018326021	NA	NA	NA
## 1993	Q3	-0.009235105	NA	NA	NA
## 1993	Q4	0.004314188	NA	NA	NA
## 1994	Q1	0.016865309	NA	NA	NA
## 1994	Q2	0.024102549	NA	NA	NA
## 1994	Q3	0.026387608	NA	NA	NA
## 1994	Q4	0.027902670	NA	NA	NA
## 1995	Q1	0.029359051	NA	NA	NA
## 1995	Q2	0.029652390	NA	NA	NA
## 1995	Q3	0.027469338	NA	NA	NA
## 1995	Q4	0.023857727	NA	NA	NA

##	1996	Q1	0.021732121	1.720959e-02	0.0199465997	0.025369485
##	1996	Q2	0.021997034	2.229846e-02	0.0226633323	0.024845384
##	1996	Q3	0.024730281	2.810018e-02	0.0281931464	0.029931357
##	1996	Q4	0.029016136	2.916869e-02	0.0269391547	0.026817261
##	1997	Q1	0.033505248	3.407855e-02	0.0343393902	0.031849824
##	1997	Q2	0.038488065	3.499521e-02	0.0349992358	0.034691827
##	1997	Q3	0.043073859	3.802733e-02	0.0374185729	0.035070421
##	1997	Q4	0.045774275	4.735475e-02	0.0482436303	0.045836288
##	1998	Q1	0.046281756	4.592731e-02	0.0443650439	0.044339165
##	1998	Q2	0.045082886	4.620195e-02	0.0457767277	0.042712325
##	1998	Q3	0.042744369	4.670864e-02	0.0468750000	0.044649837
##	1998	Q4	0.039885031	4.002706e-02	0.0417086150	0.040593405
##	1999	Q1	0.038364366	4.067039e-02	0.0417676408	0.038701338
##	1999	Q2	0.039614093	4.659657e-02	0.0482914431	0.046447213
##	1999	Q3	0.043256810	4.987422e-02	0.0497977403	0.046403978
##	1999	Q4	0.047595503	5.247181e-02	0.0494270330	0.047711543
##	2000	Q1	0.050164619	5.819197e-02	0.0567870826	0.055864943
##	2000	Q2	0.048035628	5.382826e-02	0.0513200431	0.052945860
##	2000	Q3	0.042215552	4.552558e-02	0.0451767207	0.051807505
##	2000	Q4	0.036024604	4.491141e-02	0.0473621892	0.051046578
##	2001	Q1	0.031129327	4.007711e-02	0.0401398420	0.044878578
##	2001	Q2	0.028366914	3.532363e-02	0.0360025625	0.040413626
##	2001	Q3	0.026813240	3.746513e-02	0.0390287312	0.039488024
##	2001	Q4	0.025548726	3.302445e-02	0.0317799271	0.035428745
##	2002	Q1	0.023930648	2.672910e-02	0.0277604880	0.031027712
##	2002	Q2	0.022259051	2.878744e-02	0.0299282711	0.030590611
##	2002	Q3	0.021507030	2.564349e-02	0.0234919858	0.026606461
##	2002	Q4	0.021627998	2.700639e-02	0.0271487704	0.027036747
##	2003	Q1	0.022473293	3.192399e-02	0.0305232558	0.031303229
##	2003	Q2	0.024116092	3.005464e-02	0.0294188280	0.030558264
##	2003	Q3	0.026036402	3.024628e-02	0.0308427974	0.031391113
##	2003	Q4	0.027250307	3.168556e-02	0.0327130497	0.034217668
##	2004	Q1	0.027617617	2.946322e-02	0.0326751293	0.030333689
##	2004	Q2	0.026826797	3.091558e-02	0.0296279016	0.031320950
##	2004	Q3	0.026430702	3.626613e-02	0.0372260234	0.034612468
##	2004	Q4	0.026548243	3.395479e-02	0.0307586365	0.030395810
##	2005	Q1	NA	3.568554e-02	0.0322103346	0.034531927
##	2005	Q2	NA	3.699760e-02	0.0380650278	0.037035572
##	2005	Q3	NA	3.420753e-02	0.0326475850	0.036425746
##	2005	Q4	NA	3.771777e-02	0.0405300078	0.040856298
##	2006	Q1	NA	3.920553e-02	0.0402469953	0.041595417
##	2006	Q2	NA	3.995551e-02	0.0410345957	0.041851513
##	2006	Q3	NA	4.104826e-02	0.0420095279	0.042253772
##	2006	Q4	NA	4.046000e-02	0.0397003745	0.041267240
##	2007	Q1	NA	3.922220e-02	0.0389018444	0.040646289
##	2007	Q2	NA	3.702180e-02	0.0352238180	0.038243604
##	2007	Q3	NA	3.511202e-02	0.0340814630	0.036435046
##	2007	Q4	NA	3.162566e-02	0.0310827501	0.035517777
##	2008	Q1	NA	2.750680e-02	0.0272421182	0.029708659
##	2008	Q2	NA	1.896073e-02	0.0193417722	0.021995768
##	2008	Q3	NA	2.833307e-03	0.0035168810	0.006111113
##	2008	Q4	NA	-1.423321e-02	-0.0138750250	-0.012520538
##	2009	Q1	NA	-3.470817e-02	-0.0354588796	-0.032725808
##	2009	Q2	NA	-4.445656e-02	-0.0454003576	-0.042626953
##	2009	Q3	NA	-3.924031e-02	-0.0409532392	-0.038337322
##	2009	Q4	NA	-3.038477e-02	-0.0313796943	-0.029161392
##	2010	Q1	NA	-1.378588e-02	-0.0135928329	-0.010405433
##	2010	Q2	NA	-8.147967e-05	-0.0012488292	0.001088952
##	2010	Q3	NA	2.205522e-03	0.0017749008	0.004698701
##	2010	Q4	NA	6.054656e-03	0.0051207023	0.005316482
##	2011	Q1	NA	8.910325e-03	0.0058461217	-0.001323120
##	2011	Q2	NA	7.252282e-03	0.0028133792	-0.007913307

## 2011 Q3	NA	NA	-0.0002084419	-0.012423304
## 2011 Q4	NA	NA	-0.0063422749	-0.018269449
## 2012 Q1	NA	NA	-0.0123508044	-0.023267009
## 2012 Q2	NA	NA	-0.0161055694	-0.027899857
## 2012 Q3	NA	NA	-0.0166788283	-0.031146410
## 2012 Q4	NA	NA	-0.0206131631	-0.034888870
## 2013 Q1	NA	NA	-0.0194409416	-0.029778884
## 2013 Q2	NA	NA	-0.0159467737	-0.021278978
## 2013 Q3	NA	NA	-0.0113431570	-0.014669651
## 2013 Q4	NA	NA	-0.0019230769	-0.002104142
## 2014 Q1	NA	NA	0.0048226342	0.005289577
## 2014 Q2	NA	NA	0.0119124276	0.010016015
## 2014 Q3	NA	NA		0.017584565
## 2014 Q4	NA	NA		0.022299143
## 2015 Q1	NA	NA		0.030209238
## 2015 Q2	NA	NA		0.035804326
## 2015 Q3	NA	NA		0.038796664
## 2015 Q4	NA	NA		0.040875363
## 2016 Q1	NA	NA		0.035756257
## 2016 Q2	NA	NA		0.033166445
## 2016 Q3	NA	NA		0.031031499
## 2016 Q4	NA	NA		0.027080691
## 2017 Q1	NA	NA		0.028509412
## 2017 Q2	NA	NA		0.030501459
## 2017 Q3	NA	NA		0.029240710
## 2017 Q4	NA	NA		0.030897467
## 2018 Q1	NA	NA		0.028945107
## 2018 Q2	NA	NA		0.026089611
## 2018 Q3	NA	NA		0.025031891
## 2018 Q4	NA	NA		0.023273376
## 2019 Q1	NA	NA		0.023707150
## 2019 Q2	NA	NA		0.022512449

## 4.4 PIB base 1986

```
#PIB base 1986 ----
```

```
# es un excel
library(excel.link)
```

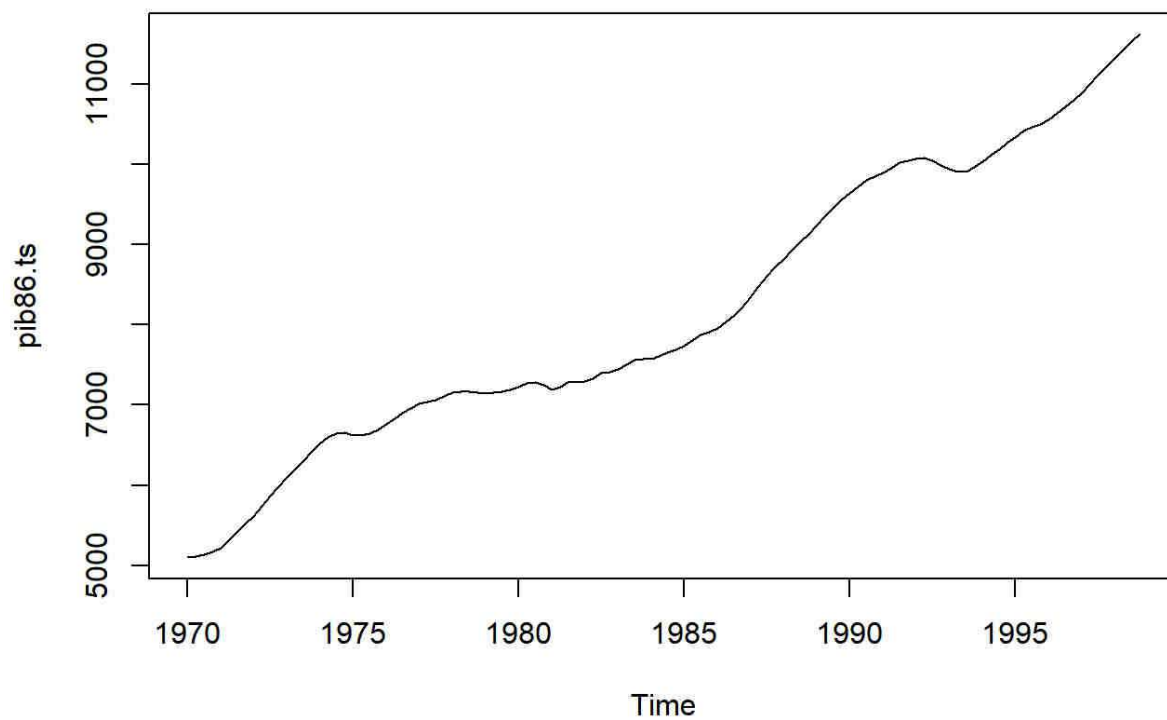
```
##
## To Daniela Khazova who constantly inspires me...
```

```
url<-"http://www.ine.es/daco/daco42/daco4214/cntrb86.xls"

datos<-xl.read.file(url, top.left.cell = "A8",header=TRUE)

pib86.ts<-ts(datos[,3], start=c(1970,1), frequency = 4)

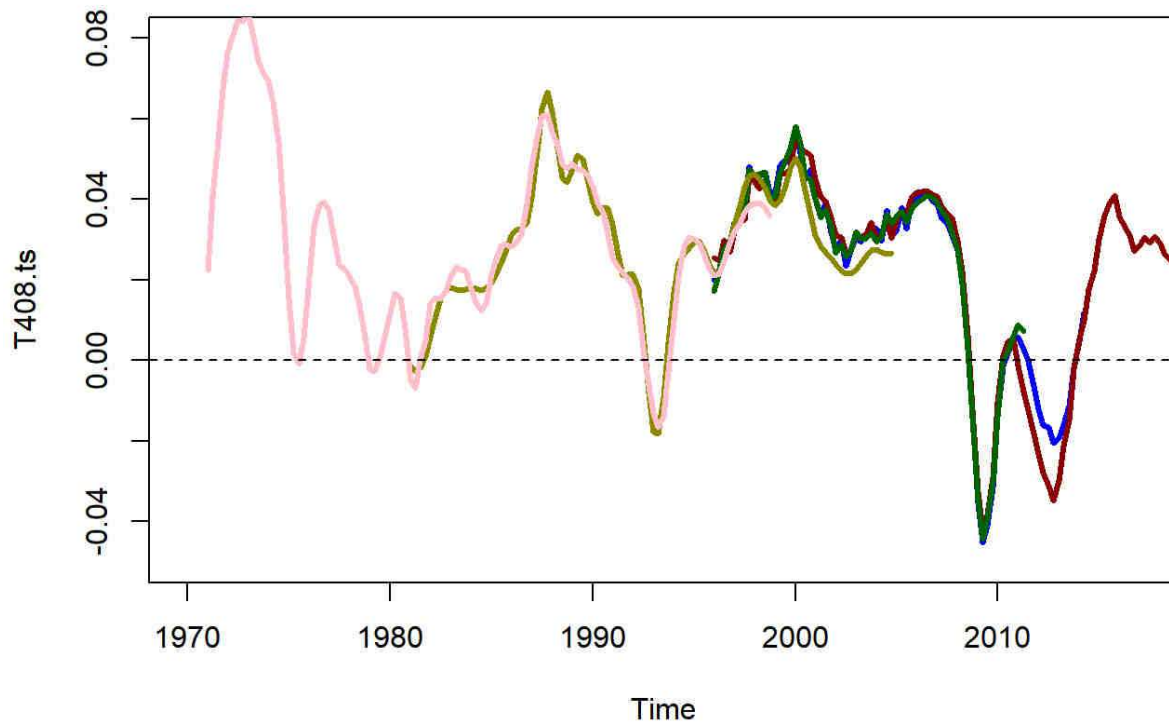
plot(pib86.ts)
```



```
T186.ts<-pib86.ts/stats::lag(pib86.ts, k = -1)-1
T486.ts<-pib86.ts/stats::lag(pib86.ts, k = -4)-1

plot(T408.ts, col="blue", xlim=c(1970,2017), ylim=c(-0.05,0.08),lwd=3)
lines(T410.ts, col="darkred",lwd=3)
lines(T400.ts, col="darkgreen",lwd=3)
lines(T495.ts, col="yellow4",lwd=3)
lines(T486.ts, col="pink",lwd=3)

abline(h=0,lty=2)
```



## 4.5 Enlaze de la serie PIB por tasas de crecimiento

```
# ahora voy a fusionar todas as series utilizando base 86, base 95, y base 2010----

#creo la variable tiempo
tiempo<-ts(seq(1970,tiempo_max,by=0.25), start = c(1970,1),frequency = 4)

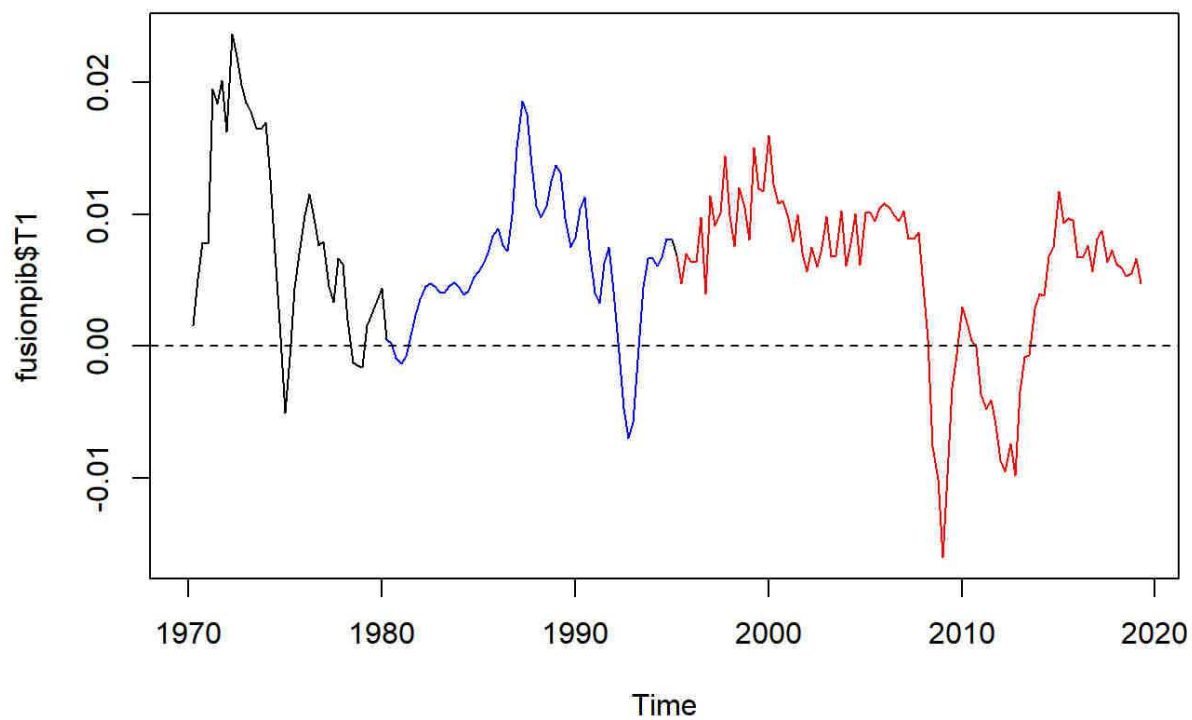
fusionpib<-ts.union(tiempo,T186.ts,T195.ts,T100.ts,T108.ts,T110.ts,pib10.ts, dframe = TRUE)

# Reconstruyo la serie de tasa de crecimiento

fusionpib$T1<-fusionpib$T110.ts
fusionpib$T1[tiempo>1980 &tiempo<=1995]<-fusionpib$T195.ts[tiempo>1980 &tiempo<=1995]
fusionpib$T1[tiempo>1970 &tiempo<=1980]<-fusionpib$T186.ts[tiempo>1970 &tiempo<=1980]

plot(fusionpib$T1)
lines(fusionpib$T110.ts, col="red")
lines(window(fusionpib$T195.ts,start=1980, end=1995), col="blue")
abline(h=0,lty=2)
```



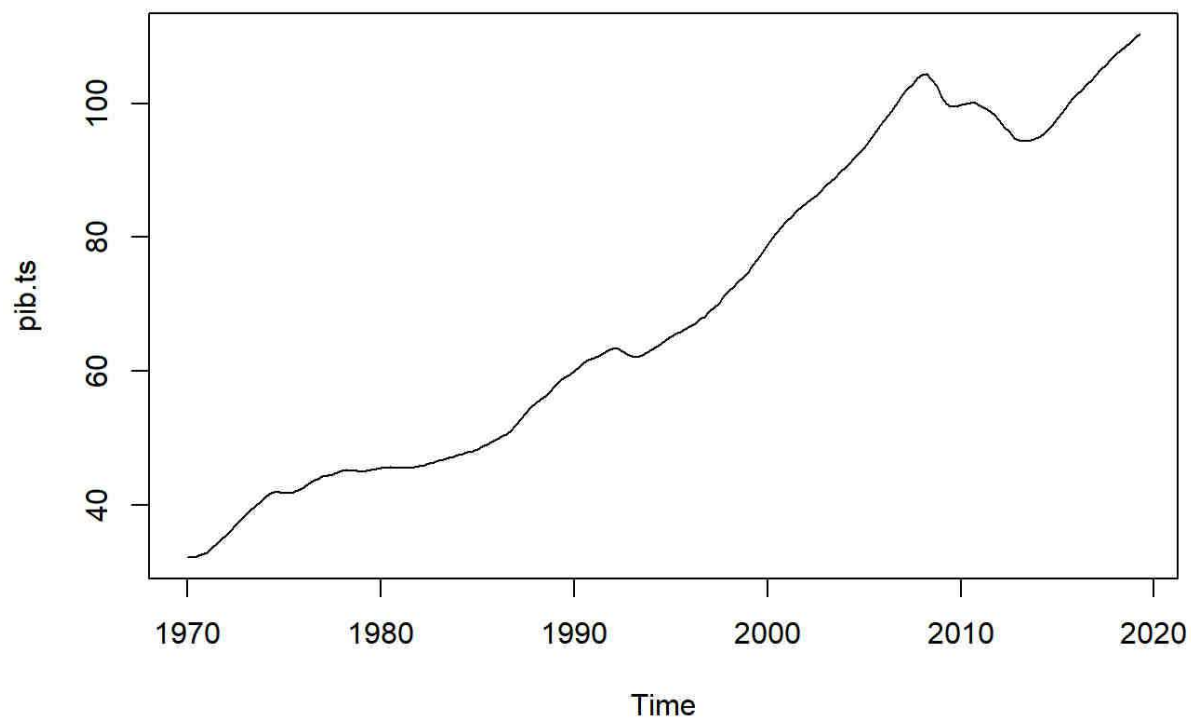


```
# Reconstruyo la serie del PIB

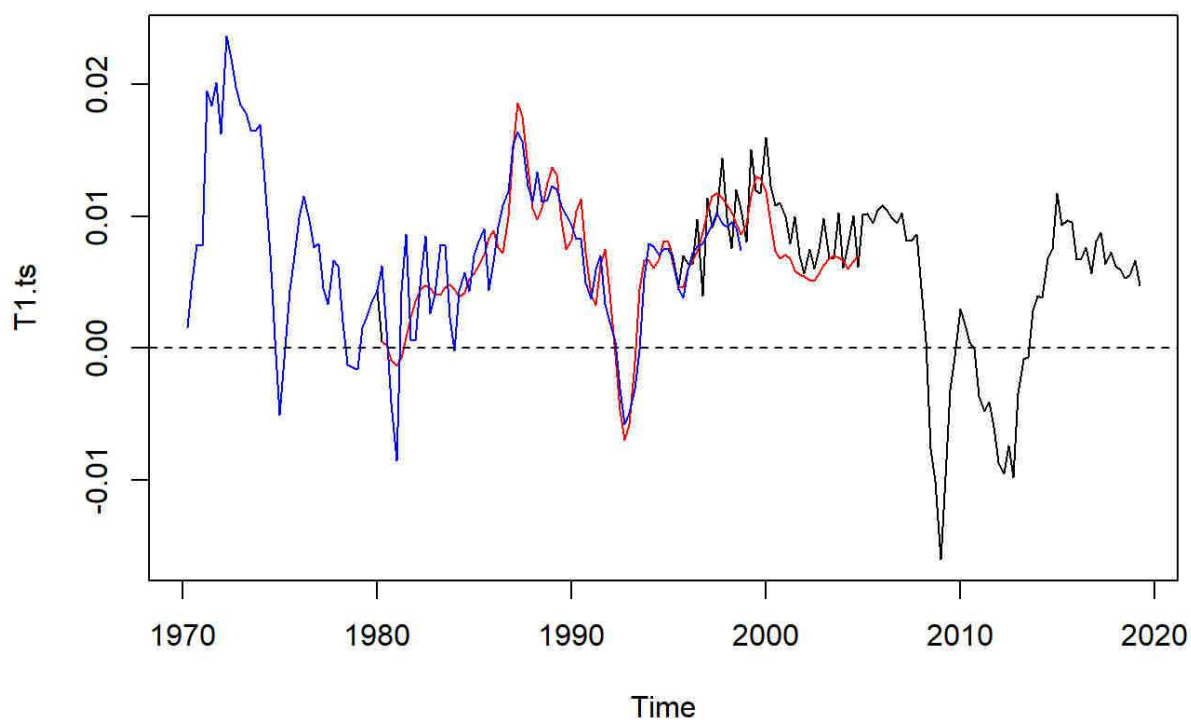
fusionpib$pib<-fusionpib$pib10.ts

for(i in seq(1994.75, 1970, by=-0.25)){
  fusionpib$pib[tiempo==i]<-fusionpib$pib[tiempo==(i+0.25)]/(1+fusionpib$T1[tiempo==(i+0.25)])
}

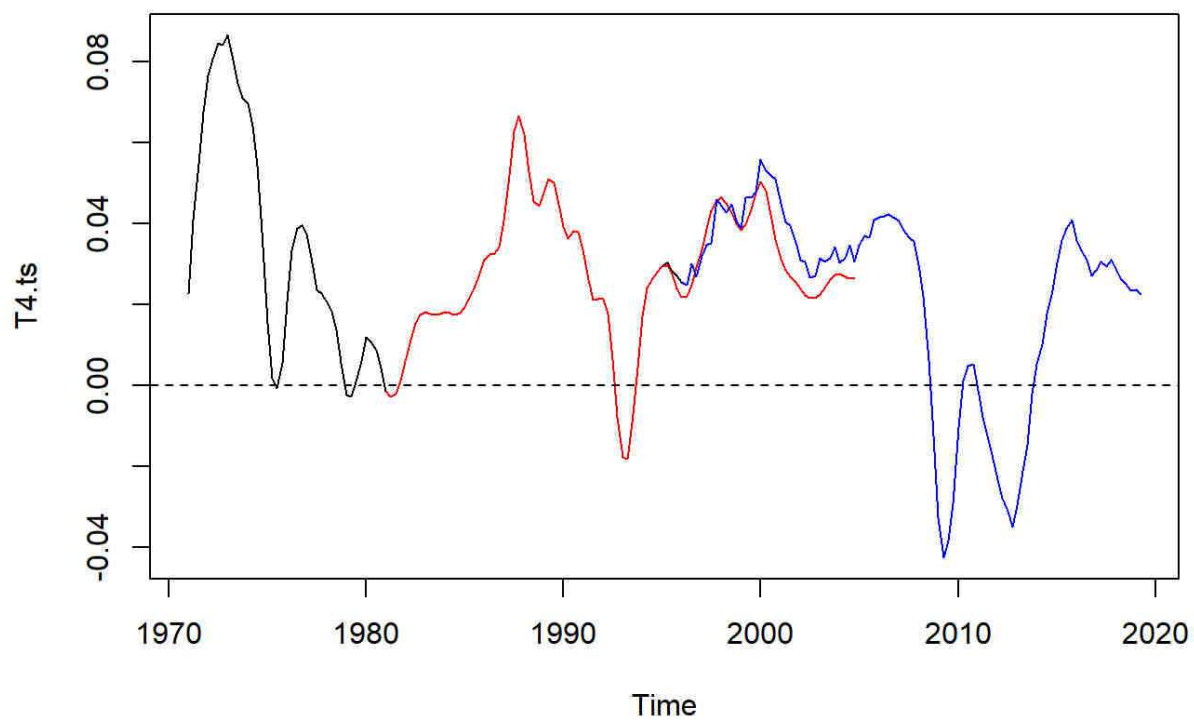
pib.ts<-fusionpib$pib
plot(pib.ts)
```



```
T1.ts<-pib.ts/stats::lag(pib.ts, k = -1)-1  
plot(T1.ts)  
lines(T195.ts,col="red")  
lines(T186.ts,col="blue")  
  
abline(h=0,lty=2)
```



```
T4.ts<-pib.ts/stats::lag(pib.ts, k = -4)-1  
plot(T4.ts)  
lines(T495.ts,col="red")  
lines(T410.ts,col="blue")  
abline(h=0,lty=2)
```



```
ts.union(time(pib.ts),pib.ts,T1.ts,T4.ts,dframe = TRUE)
```

##	time.pib.ts.	pib.ts	T1.ts	T4.ts
## 1	1970.00	32.18634	NA	NA
## 2	1970.25	32.23748	1.588640e-03	NA
## 3	1970.50	32.40224	5.110833e-03	NA
## 4	1970.75	32.65790	7.890276e-03	NA
## 5	1971.00	32.91293	7.809178e-03	0.0225743817
## 6	1971.25	33.55493	1.950593e-02	0.0408670792
## 7	1971.50	34.17231	1.839902e-02	0.0546279881
## 8	1971.75	34.86039	2.013559e-02	0.0674411412
## 9	1972.00	35.42600	1.622512e-02	0.0763550577
## 10	1972.25	36.26306	2.362836e-02	0.0807073653
## 11	1972.50	37.05972	2.196884e-02	0.0844955942
## 12	1972.75	37.79388	1.981024e-02	0.0841497202
## 13	1973.00	38.48953	1.840655e-02	0.0864769508
## 14	1973.25	39.17572	1.782786e-02	0.0803203064
## 15	1973.50	39.82277	1.651654e-02	0.0745566968
## 16	1973.75	40.47802	1.645425e-02	0.0710205445
## 17	1974.00	41.16547	1.698325e-02	0.0695237158
## 18	1974.25	41.68563	1.263590e-02	0.0640680643
## 19	1974.50	41.97980	7.056864e-03	0.0541658741
## 20	1974.75	42.00758	6.616442e-04	0.0377873429
## 21	1975.00	41.79673	-5.019160e-03	0.0153348361
## 22	1975.25	41.76201	-8.306777e-04	0.0018323616
## 23	1975.50	41.94445	4.368462e-03	-0.0008420926
## 24	1975.75	42.23925	7.028369e-03	0.0055150650
## 25	1976.00	42.65904	9.938427e-03	0.0206310130
## 26	1976.25	43.15143	1.154239e-02	0.0332698470
## 27	1976.50	43.56996	9.699080e-03	0.0387538566
## 28	1976.75	43.90453	7.678934e-03	0.0394249163
## 29	1977.00	44.25236	7.922358e-03	0.0373499859
## 30	1977.25	44.45310	4.536312e-03	0.0301651623
## 31	1977.50	44.60082	3.322967e-03	0.0236598088
## 32	1977.75	44.89877	6.680537e-03	0.0226455787
## 33	1978.00	45.17905	6.242531e-03	0.0209412134
## 34	1978.25	45.27059	2.026017e-03	0.0183899232
## 35	1978.50	45.21504	-1.227097e-03	0.0137715313
## 36	1978.75	45.14812	-1.479910e-03	0.0055536028
## 37	1979.00	45.07616	-1.593960e-03	-0.0022775224
## 38	1979.25	45.14623	1.554491e-03	-0.0027470229
## 39	1979.50	45.25481	2.405023e-03	0.0008795689
## 40	1979.75	45.41389	3.515184e-03	0.0058864653
## 41	1980.00	45.61273	4.378588e-03	0.0119037616
## 42	1980.25	45.63553	4.997698e-04	0.0108381416
## 43	1980.50	45.64573	2.234696e-04	0.0086382354
## 44	1980.75	45.60674	-8.542515e-04	0.0042464938
## 45	1981.00	45.54555	-1.341664e-03	-0.0014730059
## 46	1981.25	45.51075	-7.639319e-04	-0.0027342158
## 47	1981.50	45.54735	8.040598e-04	-0.0021553424
## 48	1981.75	45.65353	2.331217e-03	0.0010259783
## 49	1982.00	45.81490	3.534683e-03	0.0059138864
## 50	1982.25	46.02246	4.530456e-03	0.0112436565
## 51	1982.50	46.24262	4.783753e-03	0.0152648632
## 52	1982.75	46.45138	4.514497e-03	0.0174763150
## 53	1983.00	46.64395	4.145519e-03	0.0180956371
## 54	1983.25	46.83291	4.051239e-03	0.0176099481
## 55	1983.50	47.04767	4.585687e-03	0.0174093533
## 56	1983.75	47.27323	4.794267e-03	0.0176927150
## 57	1984.00	47.48440	4.466835e-03	0.0180183656
## 58	1984.25	47.67036	3.916367e-03	0.0178816175
## 59	1984.50	47.87013	4.190524e-03	0.0174812246
## 60	1984.75	48.12208	5.263290e-03	0.0179561692

## 61	1985.00	48.39383	5.647112e-03	0.0191522961
## 62	1985.25	48.69678	6.259994e-03	0.0215314919
## 63	1985.50	49.04171	7.083374e-03	0.0244742976
## 64	1985.75	49.45024	8.330173e-03	0.0275997906
## 65	1986.00	49.89116	8.916440e-03	0.0309404867
## 66	1986.25	50.27389	7.671312e-03	0.0323864196
## 67	1986.50	50.63862	7.254937e-03	0.0325622928
## 68	1986.75	51.14373	9.974767e-03	0.0342464092
## 69	1987.00	51.91279	1.503724e-02	0.0405208797
## 70	1987.25	52.87862	1.860476e-02	0.0518107511
## 71	1987.50	53.80845	1.758426e-02	0.0625969934
## 72	1987.75	54.54812	1.374628e-02	0.0665650108
## 73	1988.00	55.12761	1.062356e-02	0.0619272681
## 74	1988.25	55.66511	9.750152e-03	0.0526960645
## 75	1988.50	56.25661	1.062592e-02	0.0454976198
## 76	1988.75	56.96028	1.250826e-02	0.0442208292
## 77	1989.00	57.74194	1.372287e-02	0.0474231740
## 78	1989.25	58.50260	1.317348e-02	0.0509742219
## 79	1989.50	59.06770	9.659359e-03	0.0499690759
## 80	1989.75	59.51282	7.535749e-03	0.0448125875
## 81	1990.00	60.00233	8.225309e-03	0.0391464251
## 82	1990.25	60.62621	1.039771e-02	0.0362995017
## 83	1990.50	61.31429	1.134947e-02	0.0380342054
## 84	1990.75	61.76301	7.318338e-03	0.0378102131
## 85	1991.00	62.01196	4.030809e-03	0.0334926316
## 86	1991.25	62.21413	3.260070e-03	0.0261918426
## 87	1991.50	62.60646	6.306107e-03	0.0210744651
## 88	1991.75	63.07677	7.512241e-03	0.0212710161
## 89	1992.00	63.33952	4.165597e-03	0.0214081183
## 90	1992.25	63.30833	-4.924942e-04	0.0175876732
## 91	1992.50	63.00898	-4.728379e-03	0.0064294817
## 92	1992.75	62.57286	-6.921568e-03	-0.0079888157
## 93	1993.00	62.21113	-5.781012e-03	-0.0178150305
## 94	1993.25	62.14814	-1.012497e-03	-0.0183260212
## 95	1993.50	62.42709	4.488460e-03	-0.0092351048
## 96	1993.75	62.84281	6.659363e-03	0.0043141879
## 97	1994.00	63.26034	6.643947e-03	0.0168653090
## 98	1994.25	63.64607	6.097503e-03	0.0241025493
## 99	1994.50	64.07439	6.729754e-03	0.0263876076
## 100	1994.75	64.59629	8.145305e-03	0.0279026700
## 101	1995.00	65.11760	8.070208e-03	0.0293590510
## 102	1995.25	65.56550	6.878325e-03	0.0301579256
## 103	1995.50	65.87740	4.757075e-03	0.0281393391
## 104	1995.75	66.34160	7.046423e-03	0.0270186634
## 105	1996.00	66.76960	6.451457e-03	0.0253694854
## 106	1996.25	67.19450	6.363674e-03	0.0248453836
## 107	1996.50	67.84920	9.743357e-03	0.0299313573
## 108	1996.75	68.12070	4.001521e-03	0.0268172610
## 109	1997.00	68.89620	1.138420e-02	0.0318498239
## 110	1997.25	69.52560	9.135482e-03	0.0346918275
## 111	1997.50	70.22870	1.011282e-02	0.0350704209
## 112	1997.75	71.24310	1.444424e-02	0.0458362876
## 113	1998.00	71.95100	9.936401e-03	0.0443391653
## 114	1998.25	72.49520	7.563481e-03	0.0427123247
## 115	1998.50	73.36440	1.198976e-02	0.0446498369
## 116	1998.75	74.13510	1.050510e-02	0.0405934048
## 117	1999.00	74.73560	8.100077e-03	0.0387013384
## 118	1999.25	75.86240	1.507715e-02	0.0464472131
## 119	1999.50	76.76880	1.194795e-02	0.0464039780
## 120	1999.75	77.67220	1.176780e-02	0.0477115428
## 121	2000.00	78.91070	1.594522e-02	0.0558649425
## 122	2000.25	79.87900	1.227083e-02	0.0529458599

## 123	2000.50	80.74600	1.085392e-02	0.0518075051
## 124	2000.75	81.63710	1.103584e-02	0.0510465778
## 125	2001.00	82.45210	9.983206e-03	0.0448785779
## 126	2001.25	83.10720	7.945219e-03	0.0404136256
## 127	2001.50	83.93450	9.954613e-03	0.0394880242
## 128	2001.75	84.52940	7.087670e-03	0.0354287450
## 129	2002.00	85.01040	5.690328e-03	0.0310277118
## 130	2002.25	85.64950	7.517904e-03	0.0305906107
## 131	2002.50	86.16770	6.050240e-03	0.0266064610
## 132	2002.75	86.81480	7.509775e-03	0.0270367470
## 133	2003.00	87.67150	9.868133e-03	0.0313032288
## 134	2003.25	88.26680	6.790120e-03	0.0305582636
## 135	2003.50	88.87260	6.863283e-03	0.0313911129
## 136	2003.75	89.78540	1.027088e-02	0.0342176680
## 137	2004.00	90.33090	6.075598e-03	0.0303336888
## 138	2004.25	91.03140	7.754821e-03	0.0313209497
## 139	2004.50	91.94870	1.007674e-02	0.0346124677
## 140	2004.75	92.51450	6.153431e-03	0.0303958105
## 141	2005.00	93.45020	1.011409e-02	0.0345319265
## 142	2005.25	94.40280	1.019366e-02	0.0370355723
## 143	2005.50	95.29800	9.482770e-03	0.0364257461
## 144	2005.75	96.29430	1.045457e-02	0.0408562982
## 145	2006.00	97.33730	1.083138e-02	0.0415954166
## 146	2006.25	98.35370	1.044204e-02	0.0418515129
## 147	2006.50	99.32470	9.872531e-03	0.0422537724
## 148	2006.75	100.26810	9.498141e-03	0.0412672401
## 149	2007.00	101.29370	1.022858e-02	0.0406462887
## 150	2007.25	102.11510	8.109093e-03	0.0382436045
## 151	2007.50	102.94360	8.113394e-03	0.0364350459
## 152	2007.75	103.82940	8.604712e-03	0.0355177768
## 153	2008.00	104.30300	4.561328e-03	0.0297086591
## 154	2008.25	104.36120	5.579897e-04	0.0219957675
## 155	2008.50	103.57270	-7.555490e-03	0.0061111133
## 156	2008.75	102.52940	-1.007312e-02	-0.0125205385
## 157	2009.00	100.88960	-1.599346e-02	-0.0327258085
## 158	2009.25	99.91260	-9.683852e-03	-0.0426269533
## 159	2009.50	99.60200	-3.108717e-03	-0.0383373225
## 160	2009.75	99.53950	-6.274974e-04	-0.0291613918
## 161	2010.00	99.83980	3.016893e-03	-0.0104054333
## 162	2010.25	100.02140	1.818914e-03	0.0010889517
## 163	2010.50	100.07000	4.858960e-04	0.0046987008
## 164	2010.75	100.06870	-1.299091e-05	0.0053164824
## 165	2011.00	99.70770	-3.607522e-03	-0.0013231196
## 166	2011.25	99.22990	-4.792007e-03	-0.0079133066
## 167	2011.50	98.82680	-4.062284e-03	-0.0124233037
## 168	2011.75	98.24050	-5.932601e-03	-0.0182694489
## 169	2012.00	97.38780	-8.679720e-03	-0.0232670095
## 170	2012.25	96.46140	-9.512485e-03	-0.0278998568
## 171	2012.50	95.74870	-7.388448e-03	-0.0311464097
## 172	2012.75	94.81300	-9.772456e-03	-0.0348888697
## 173	2013.00	94.48770	-3.430964e-03	-0.0297788840
## 174	2013.25	94.40880	-8.350293e-04	-0.0212789779
## 175	2013.50	94.34410	-6.853175e-04	-0.0146696509
## 176	2013.75	94.61350	2.855504e-03	-0.0021041418
## 177	2014.00	94.98750	3.952924e-03	0.0052895774
## 178	2014.25	95.35440	3.862614e-03	0.0100160155
## 179	2014.50	96.00310	6.803042e-03	0.0175845654
## 180	2014.75	96.72330	7.501841e-03	0.0222991434
## 181	2015.00	97.85700	1.172106e-02	0.0302092381
## 182	2015.25	98.76850	9.314612e-03	0.0358043258
## 183	2015.50	99.72770	9.711598e-03	0.0387966639
## 184	2015.75	100.67690	9.517917e-03	0.0408753630

```
## 185      2016.00 101.35600 6.745341e-03 0.0357562566
## 186      2016.25 102.04430 6.790915e-03 0.0331664448
## 187      2016.50 102.82240 7.625120e-03 0.0310314988
## 188      2016.75 103.40330 5.649547e-03 0.0270806908
## 189      2017.00 104.24560 8.145775e-03 0.0285094124
## 190      2017.25 105.15680 8.740896e-03 0.0305014587
## 191      2017.50 105.82900 6.392359e-03 0.0292407102
## 192      2017.75 106.59820 7.268329e-03 0.0308974665
## 193      2018.00 107.26300 6.236503e-03 0.0289451066
## 194      2018.25 107.90030 5.941471e-03 0.0260896109
## 195      2018.50 108.47810 5.354943e-03 0.0250318911
## 196      2018.75 109.07910 5.540289e-03 0.0232733761
## 197      2019.00 109.80590 6.663055e-03 0.0237071497
## 198      2019.25 110.32940 4.767503e-03 0.0225124490
```

## 5 Descarga de los datos de Ocupados (Contabilidad Nacional \_Trimestral)

```
# Bajo los datos de empleo -----
```

### 5.1 Datos de Empleo (Base 2010=100)

```
## Datos de Empleo (Base 2010=100) #####

#Ocupados/puestos de trabajoequivalentes a tiempo completo

#url<-"http://www.ine.es/jaxiT3/files/t/es/px/9377.px?nocab=1"

url<-"https://www.ine.es/jaxiT3/files/t/es/px/28607.px?nocab=1"

#datos<-read.px(url) # es una lista

datos<-as.data.frame(read.px(url), stringsAsFactor=FALSE) #como un data.frame

names(datos)
```

```
## [1] "Periodo"          "Niveles.y.Tasas"  "Ramas.de.actividad"
## [4] "Tipo.de.dato"     "Población"        "Situación.profesional"
## [7] "value"
```

```
table(datos$Periodo)
```

```
##
## 2019T2 2019T1 2018T4 2018T3 2018T2 2018T1 2017T4 2017T3 2017T2 2017T1
##      672      672      672      672      672      672      672      672      672      672
## 2016T4 2016T3 2016T2 2016T1 2015T4 2015T3 2015T2 2015T1 2014T4 2014T3
##      672      672      672      672      672      672      672      672      672      672
## 2014T2 2014T1 2013T4 2013T3 2013T2 2013T1 2012T4 2012T3 2012T2 2012T1
##      672      672      672      672      672      672      672      672      672      672
## 2011T4 2011T3 2011T2 2011T1 2010T4 2010T3 2010T2 2010T1 2009T4 2009T3
##      672      672      672      672      672      672      672      672      672      672
## 2009T2 2009T1 2008T4 2008T3 2008T2 2008T1 2007T4 2007T3 2007T2 2007T1
##      672      672      672      672      672      672      672      672      672      672
## 2006T4 2006T3 2006T2 2006T1 2005T4 2005T3 2005T2 2005T1 2004T4 2004T3
##      672      672      672      672      672      672      672      672      672      672
## 2004T2 2004T1 2003T4 2003T3 2003T2 2003T1 2002T4 2002T3 2002T2 2002T1
##      672      672      672      672      672      672      672      672      672      672
## 2001T4 2001T3 2001T2 2001T1 2000T4 2000T3 2000T2 2000T1 1999T4 1999T3
##      672      672      672      672      672      672      672      672      672      672
## 1999T2 1999T1 1998T4 1998T3 1998T2 1998T1 1997T4 1997T3 1997T2 1997T1
##      672      672      672      672      672      672      672      672      672      672
## 1996T4 1996T3 1996T2 1996T1 1995T4 1995T3 1995T2 1995T1
##      672      672      672      672      672      672      672      672
```

```
table(datos$Niveles.y.Tasas)
```

```
##
##      Dato base Variación trimestral      Variación anual
##      21952      21952      21952
```

```
table(datos$Ramas.de.actividad)
```



```
##
##                                     Total
##                                     4704
##      A Agricultura, ganadería, silvicultura y pesca
##                                     4704
##      Industria (B-E, CNAE 2009)
##                                     4704
##      C Industria manufacturera
##                                     4704
##      F Construcción
##                                     4704
##      Servicios (G-T, CNAE 2009)
##                                     4704
##      G-I Comercio, transporte y hostelería
##                                     4704
##      J Información y comunicaciones
##                                     4704
##      K Actividades financieras y de seguros
##                                     4704
##      L Actividades inmobiliarias
##                                     4704
## M-N Actividades profesionales, científicas y técnicas y otras
##                                     4704
##      O-Q Administración pública, educación y sanidad
##                                     4704
##      R-T Actividades artísticas, recreativas y otros servicios
##                                     4704
##      Total Nacional
##                                     4704
```

```
table(datos$Tipo.de.dato)
```

```
##
## Datos no ajustados de estacionalidad y calendario
##                                     32928
##      Datos ajustados de estacionalidad y calendario
##                                     32928
```

```
table(datos$Población)
```

```
##
##                                     Personas
##                                     16464
##      Puestos de trabajo
##                                     16464
## Puestos de trabajo equivalentes a tiempo completo
##                                     16464
##      Horas trabajadas
##                                     16464
```

```
table(datos$Situación.profesional)
```

```
##
##      Ocupados Asalariados
##      32928      32928
```

```

ocu<-datos%>%
  dplyr::filter(Niveles.y.Tasas == "Dato base")%>%
  dplyr::filter(Ramas.de.actividad=="Total")%>%
  dplyr::filter(Tipo.de.dato=="Datos ajustados de estacionalidad y calendario")%>%
  dplyr::filter(Población=="Puestos de trabajo equivalentes a tiempo completo")%>%
  dplyr::filter(Situación.profesional=="Ocupados")%>%
  dplyr::mutate(ANO=as.numeric(substr(Periodo,1,4)))%>%
  dplyr::mutate(TRIM=as.numeric(substr(Periodo,6,7)))%>%
  dplyr::mutate(tiempo=ANO+TRIM*0.25-0.25)%>%
  dplyr::arrange(tiempo)%>%
  dplyr::select(Periodo, ANO, TRIM, tiempo, value)

head(ocu)

```

```

##   Periodo ANO TRIM tiempo  value
## 1  1995T1 1995    1 1995.00 13241.0
## 2  1995T2 1995    2 1995.25 13292.3
## 3  1995T3 1995    3 1995.50 13282.9
## 4  1995T4 1995    4 1995.75 13323.7
## 5  1996T1 1996    1 1996.00 13305.3
## 6  1996T2 1996    2 1996.25 13366.0

```

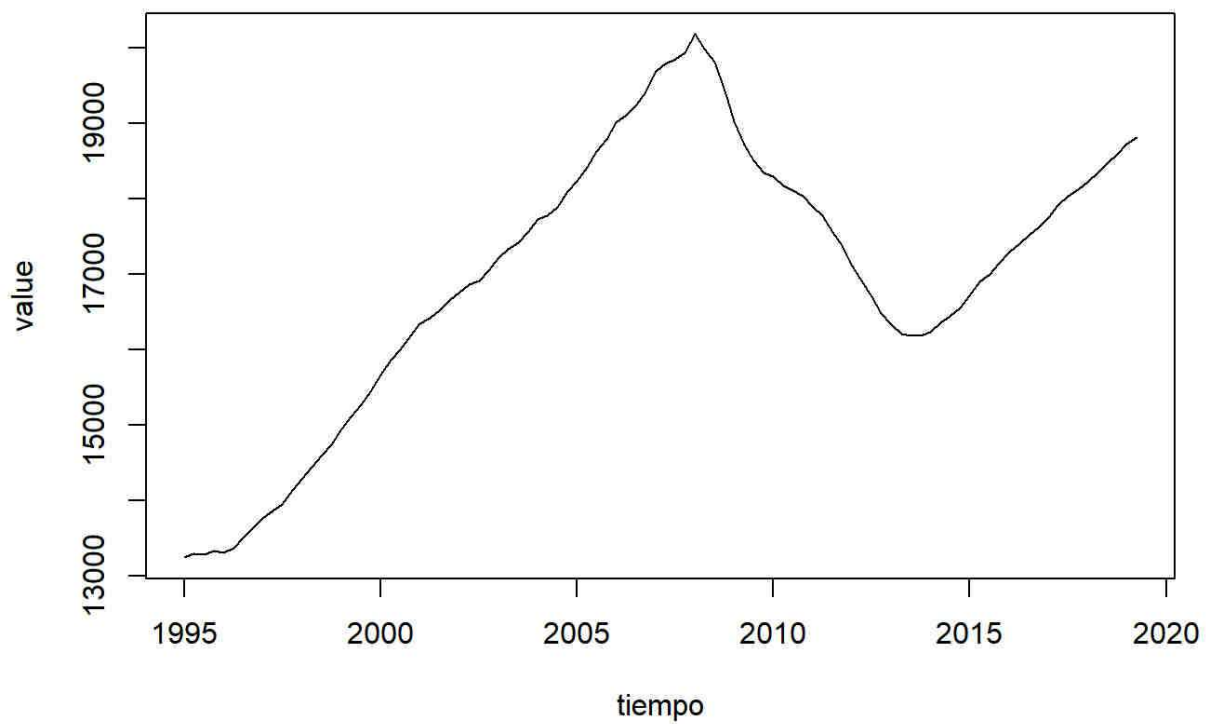
```
tail(ocu)
```

```

##   Periodo ANO TRIM tiempo  value
## 93  2018T1 2018    1 2018.00 18215.95
## 94  2018T2 2018    2 2018.25 18346.26
## 95  2018T3 2018    3 2018.50 18467.90
## 96  2018T4 2018    4 2018.75 18587.88
## 97  2019T1 2019    1 2019.00 18726.17
## 98  2019T2 2019    2 2019.25 18805.48

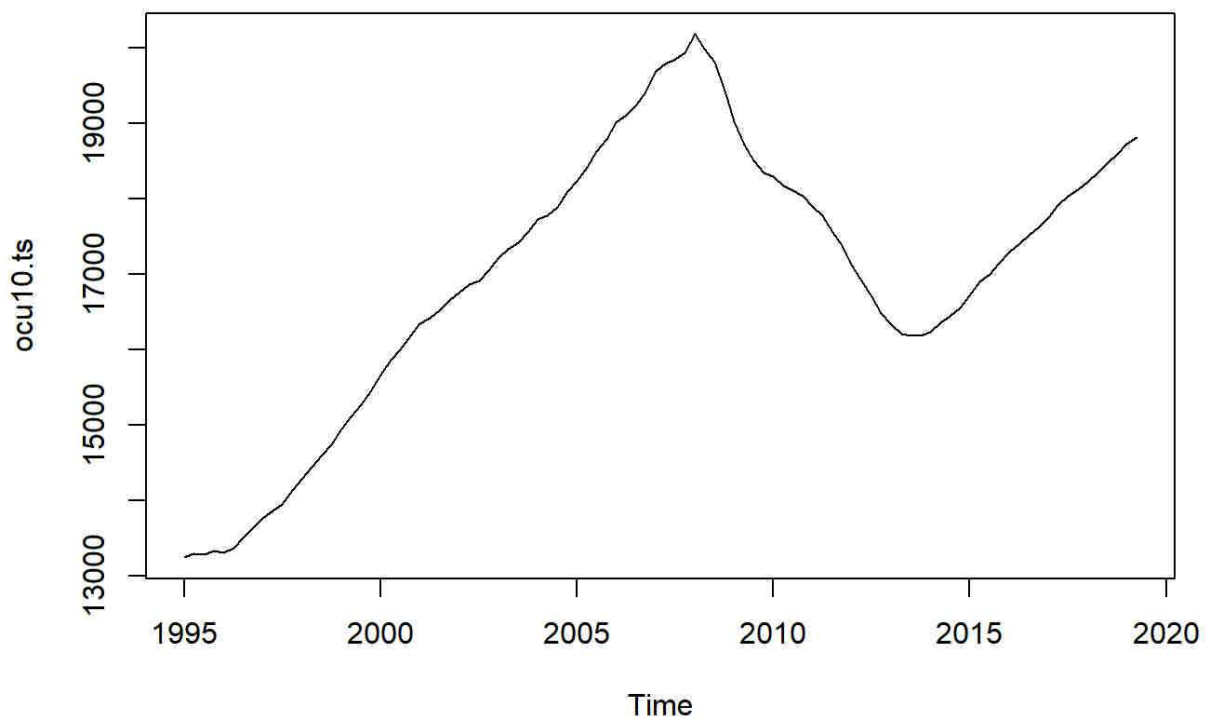
```

```
plot(value~tiempo,data = ocu,type="l")
```

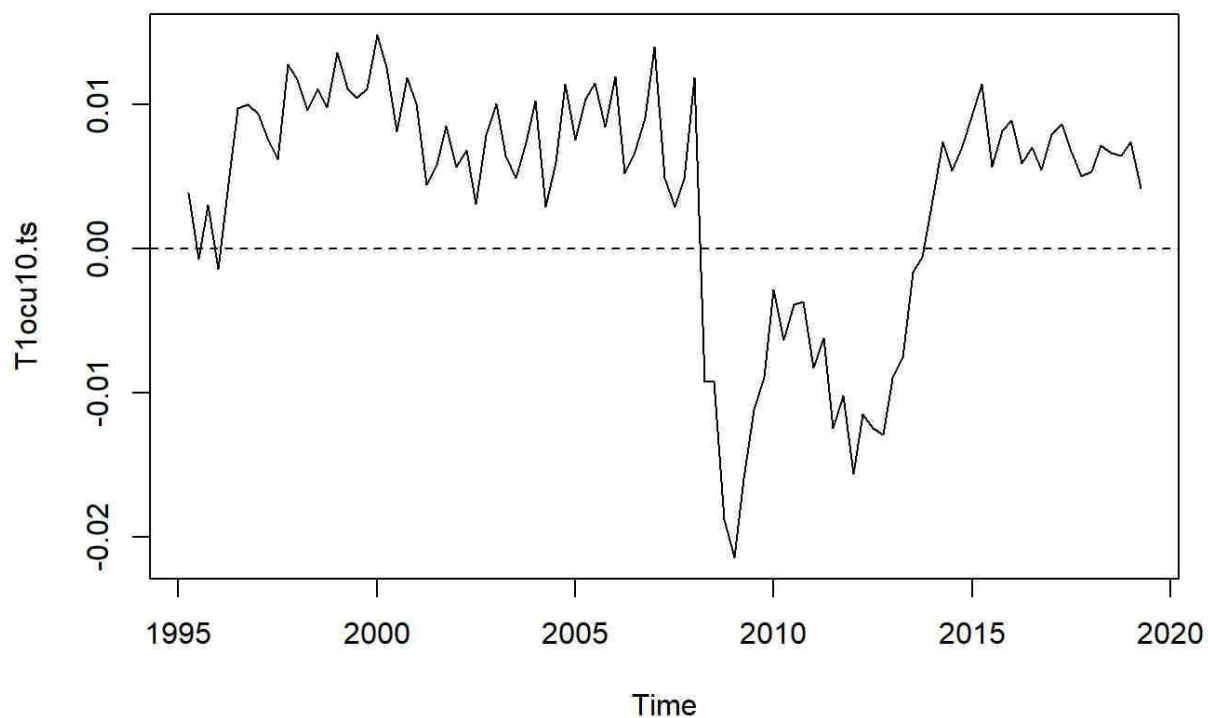


```
#Puedo convertirlo en objeto de serie temporal
ocu10.ts<-ts(ocu$value, start=c(1995,1), frequency = 4)

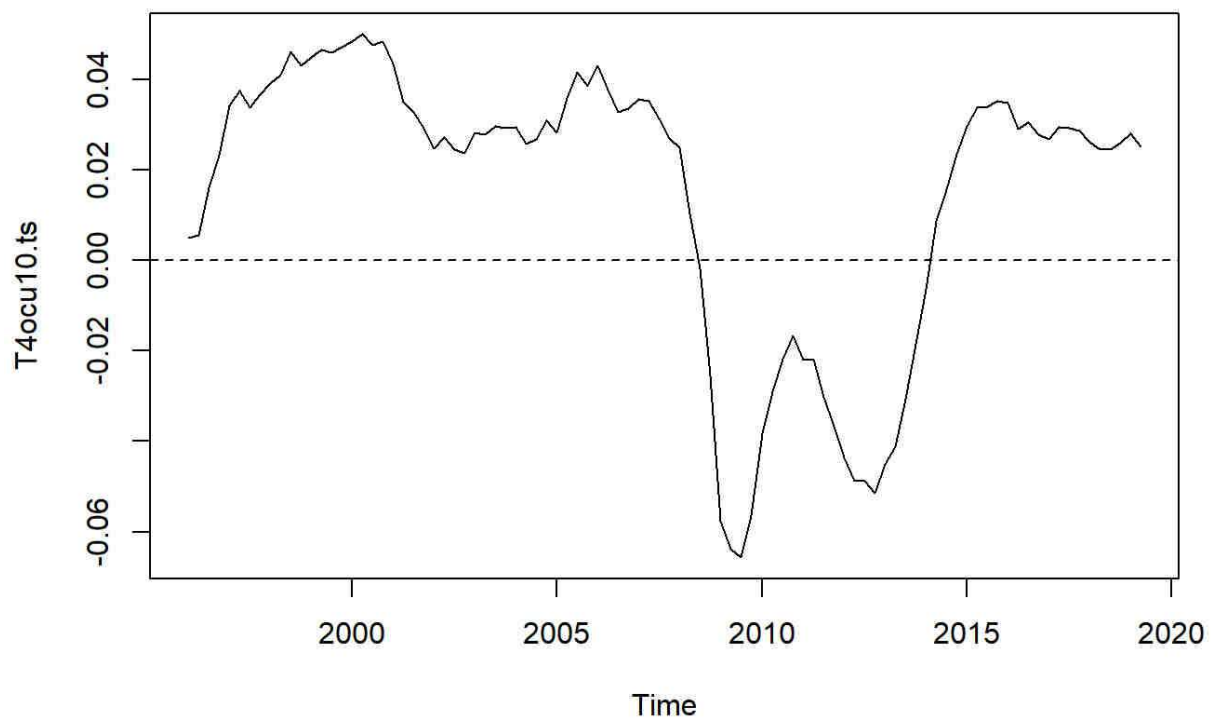
plot(ocu10.ts)
```



```
T1ocu10.ts<-ocu10.ts/stats::lag(ocu10.ts, k = -1)-1  
plot(T1ocu10.ts)  
abline(h=0,lty=2)
```



```
T4ocu10.ts<-ocu10.ts/stats::lag(ocu10.ts, k = -4)-1  
plot(T4ocu10.ts)  
abline(h=0,lty=2)
```



```
ts.union(ocu10.ts,T1ocu10.ts,T4ocu10.ts, dframe = TRUE)
```

##	ocu10.ts	T1ocu10.ts	T4ocu10.ts
## 1	13241.00	NA	NA
## 2	13292.30	0.0038743297	NA
## 3	13282.90	-0.0007071763	NA
## 4	13323.70	0.0030716184	NA
## 5	13305.30	-0.0013809978	0.004856129
## 6	13366.00	0.0045620918	0.005544563
## 7	13496.90	0.0097935059	0.016110940
## 8	13632.50	0.0100467515	0.023176745
## 9	13760.30	0.0093746562	0.034196899
## 10	13866.40	0.0077105877	0.037438276
## 11	13952.60	0.0062164657	0.033763309
## 12	14131.50	0.0128219830	0.036603704
## 13	14298.00	0.0117821887	0.039076183
## 14	14435.40	0.0096097356	0.041034443
## 15	14595.80	0.0111115729	0.046098935
## 16	14739.50	0.0098452980	0.043024449
## 17	14940.70	0.0136503952	0.044950343
## 18	15106.80	0.0111172837	0.046510661
## 19	15264.80	0.0104588662	0.045835103
## 20	15434.40	0.0111105288	0.047145426
## 21	15663.40	0.0148369875	0.048371228
## 22	15860.90	0.0126090121	0.049917918
## 23	15990.10	0.0081458177	0.047514543
## 24	16180.10	0.0118823522	0.048314155
## 25	16343.60	0.0101050055	0.043426076
## 26	16417.10	0.0044971732	0.035067367
## 27	16512.40	0.0058049229	0.032663961
## 28	16653.40	0.0085390373	0.029251982
## 29	16748.10	0.0056865265	0.024749749
## 30	16862.60	0.0068365964	0.027136340
## 31	16915.40	0.0031311897	0.024405901
## 32	17048.40	0.0078626577	0.023718880
## 33	17220.70	0.0101065203	0.028218126
## 34	17332.40	0.0064863798	0.027860472
## 35	17417.90	0.0049329579	0.029706658
## 36	17546.00	0.0073545031	0.029187490
## 37	17726.00	0.0102587484	0.029342594
## 38	17777.90	0.0029279025	0.025703307
## 39	17884.30	0.0059849589	0.026777051
## 40	18089.20	0.0114569762	0.030958623
## 41	18226.10	0.0075680517	0.028212795
## 42	18415.40	0.0103862044	0.035859128
## 43	18627.40	0.0115121040	0.041550410
## 44	18784.70	0.0084445494	0.038448356
## 45	19009.20	0.0119512156	0.042965857
## 46	19109.40	0.0052711319	0.037685850
## 47	19238.30	0.0067453714	0.032795774
## 48	19413.50	0.0091068338	0.033474051
## 49	19686.30	0.0140520772	0.035619595
## 50	19783.30	0.0049272845	0.035265367
## 51	19841.60	0.0029469300	0.031359320
## 52	19938.50	0.0048836787	0.027043037
## 53	20175.10	0.0118664895	0.024829450
## 54	19988.70	-0.0092391116	0.010382494
## 55	19803.90	-0.0092452236	-0.001900048
## 56	19431.60	-0.0187993274	-0.025423176
## 57	19014.90	-0.0214444513	-0.057506530
## 58	18711.90	-0.0159348721	-0.063876090
## 59	18503.40	-0.0111426418	-0.065668883
## 60	18338.20	-0.0089280889	-0.056269170

```
## 61 18286.10 -0.0028410640 -0.038327838
## 62 18170.80 -0.0063053357 -0.028917427
## 63 18101.00 -0.0038413278 -0.021747355
## 64 18034.40 -0.0036793547 -0.016566511
## 65 17885.40 -0.0082619882 -0.021912819
## 66 17775.00 -0.0061726324 -0.021782200
## 67 17553.40 -0.0124669480 -0.030252472
## 68 17374.90 -0.0101689701 -0.036569001
## 69 17103.71 -0.0156080898 -0.043705425
## 70 16907.10 -0.0114955170 -0.048827286
## 71 16696.96 -0.0124287466 -0.048790491
## 72 16481.13 -0.0129264840 -0.051440411
## 73 16333.24 -0.0089732936 -0.045047125
## 74 16210.47 -0.0075165745 -0.041203235
## 75 16184.64 -0.0015935999 -0.030683787
## 76 16174.91 -0.0006010021 -0.018580039
## 77 16229.45 0.0033720748 -0.006354343
## 78 16349.74 0.0074115261 0.008591239
## 79 16439.06 0.0054633910 0.015720219
## 80 16555.60 0.0070894560 0.023536270
## 81 16710.73 0.0093699385 0.029654669
## 82 16902.33 0.0114658067 0.033798466
## 83 16998.61 0.0056960779 0.034037711
## 84 17138.29 0.0082169665 0.035195392
## 85 17290.98 0.0089093507 0.034723019
## 86 17394.23 0.0059713225 0.029102197
## 87 17516.82 0.0070481430 0.030485730
## 88 17612.97 0.0054890087 0.027697519
## 89 17752.95 0.0079472666 0.026717519
## 90 17906.29 0.0086372692 0.029438445
## 91 18027.35 0.0067610894 0.029145009
## 92 18118.42 0.0050516571 0.028697368
## 93 18215.95 0.0053831405 0.026080456
## 94 18346.26 0.0071535095 0.024571036
## 95 18467.90 0.0066303425 0.024437976
## 96 18587.88 0.0064965142 0.025910703
## 97 18726.17 0.0074397937 0.028009348
## 98 18805.48 0.0042350356 0.025030439
```

## 5.2 Datos de Empleo (Base 1995=100)

```
## Datos de Empleo (Base 1995=100) #####

#Ocupados/puestos de trabajoequivalentes a tiempo completo

url<-"http://www.ine.es/jaxiT3/files/t/es/px/3160.px?nocab=1"

#datos<-read.px(url) # es una lista

datos<-as.data.frame(read.px(url), stringsAsFactor=FALSE) #como un data.frame

names(datos)
```

```
## [1] "Periodo" "Niveles.y.Tasas"
## [3] "Empleo.por.Rama.de.actividad" "Datos"
## [5] "Población" "value"
```

table(datos\$Periodo)

```
##
## 2004T4 2004T3 2004T2 2004T1 2003T4 2003T3 2003T2 2003T1 2002T4 2002T3
##      405      405      405      405      405      405      405      405      405      405
## 2002T2 2002T1 2001T4 2001T3 2001T2 2001T1 2000T4 2000T3 2000T2 2000T1
##      405      405      405      405      405      405      405      405      405      405
## 1999T4 1999T3 1999T2 1999T1 1998T4 1998T3 1998T2 1998T1 1997T4 1997T3
##      405      405      405      405      405      405      405      405      405      405
## 1997T2 1997T1 1996T4 1996T3 1996T2 1996T1 1995T4 1995T3 1995T2 1995T1
##      405      405      405      405      405      405      405      405      405      405
## 1994T4 1994T3 1994T2 1994T1 1993T4 1993T3 1993T2 1993T1 1992T4 1992T3
##      405      405      405      405      405      405      405      405      405      405
## 1992T2 1992T1 1991T4 1991T3 1991T2 1991T1 1990T4 1990T3 1990T2 1990T1
##      405      405      405      405      405      405      405      405      405      405
## 1989T4 1989T3 1989T2 1989T1 1988T4 1988T3 1988T2 1988T1 1987T4 1987T3
##      405      405      405      405      405      405      405      405      405      405
## 1987T2 1987T1 1986T4 1986T3 1986T2 1986T1 1985T4 1985T3 1985T2 1985T1
##      405      405      405      405      405      405      405      405      405      405
## 1984T4 1984T3 1984T2 1984T1 1983T4 1983T3 1983T2 1983T1 1982T4 1982T3
##      405      405      405      405      405      405      405      405      405      405
## 1982T2 1982T1 1981T4 1981T3 1981T2 1981T1 1980T4 1980T3 1980T2 1980T1
##      405      405      405      405      405      405      405      405      405      405
```

table(datos\$Niveles.y.Tasas)

```
##
##      Dato base Variación trimestral      Variación anual
##      13500      13500      13500
```

table(datos\$Empleo.por.Rama.de.actividad)



```

##
##                                Ocupados. Total ramas
##                                2700
##  Ocupados. Ramas de la agricultura, ganadería, silvicultura y pesca
##                                2700
##                                Ocupados. Ramas energéticas
##                                2700
##                                Ocupados. Ramas de la industria
##                                2700
##                                Ocupados. Ramas de la construcción
##                                2700
##                                Ocupados. Ramas de los servicios de mercado
##                                2700
##                                Ocupados. Ramas de los servicios no de mercado
##                                2700
##                                Asalariados. Total ramas
##                                2700
##  Asalariados. Ramas de la agricultura, ganadería, silvicultura y pesca
##                                2700
##                                Asalariados. Ramas energéticas
##                                2700
##                                Asalariados. Ramas de la industria
##                                2700
##                                Asalariados. Ramas de la construcción
##                                2700
##                                Asalariados. Ramas de los servicios de mercado
##                                2700
##                                Asalariados. Ramas de los servicios no de mercado
##                                2700
##                                No Asalariados
##                                2700

```

```
table(datos$Datos)
```

```

##
##                                Datos brutos
##                                13500
##  Datos corregidos de estacionalidad y calendario
##                                13500
##                                Ciclotendencia
##                                13500

```

```
table(datos$Población)
```

```

##
##                                Personas
##                                13500
##                                Puestos de trabajo
##                                13500
##  Puestos de trabajo equivalentes a tiempo completo
##                                13500

```

```

ocu<-datos%>%
  dplyr::filter(Niveles.y.Tasas == "Dato base")%>%
  dplyr::filter(Empleo.por.Rama.de.actividad=="Ocupados. Total ramas")%>%
  dplyr::filter(Datos=="Ciclotendencia")%>%
  dplyr::filter(Población=="Puestos de trabajo equivalentes a tiempo completo")%>%
  dplyr::mutate(ANO=as.numeric(substr(Periodo,1,4)))%>%
  dplyr::mutate(TRIM=as.numeric(substr(Periodo,6,7)))%>%
  dplyr::mutate(tiempo=ANO+TRIM*0.25-0.25)%>%
  dplyr::arrange(tiempo)%>%
  dplyr::select(Periodo, ANO, TRIM, tiempo, value)

head(ocu)

```

```

##   Periodo  ANO TRIM  tiempo  value
## 1  1980T1 1980    1 1980.00 12114.4
## 2  1980T2 1980    2 1980.25 12021.1
## 3  1980T3 1980    3 1980.50 11928.8
## 4  1980T4 1980    4 1980.75 11848.0
## 5  1981T1 1981    1 1981.00 11770.8
## 6  1981T2 1981    2 1981.25 11697.4

```

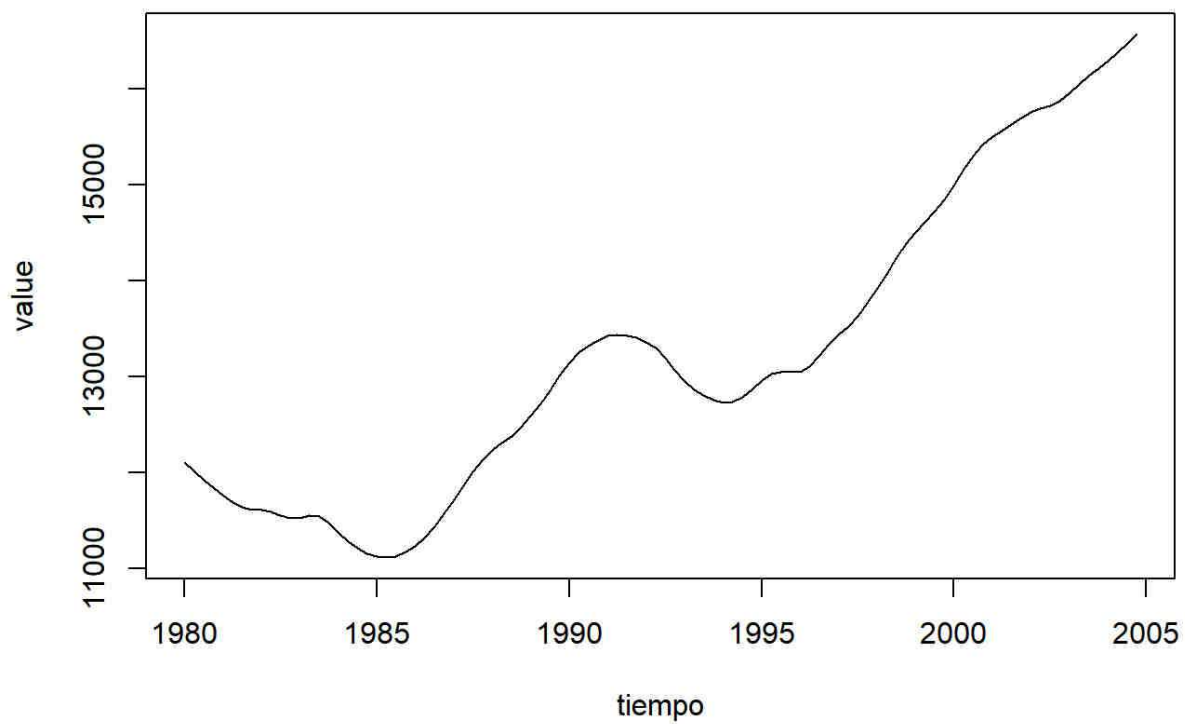
```
tail(ocu)
```

```

##   Periodo  ANO TRIM  tiempo  value
## 95  2003T3 2003    3 2003.50 16140.3
## 96  2003T4 2003    4 2003.75 16212.2
## 97  2004T1 2004    1 2004.00 16291.5
## 98  2004T2 2004    2 2004.25 16383.1
## 99  2004T3 2004    3 2004.50 16481.0
## 100 2004T4 2004    4 2004.75 16572.2

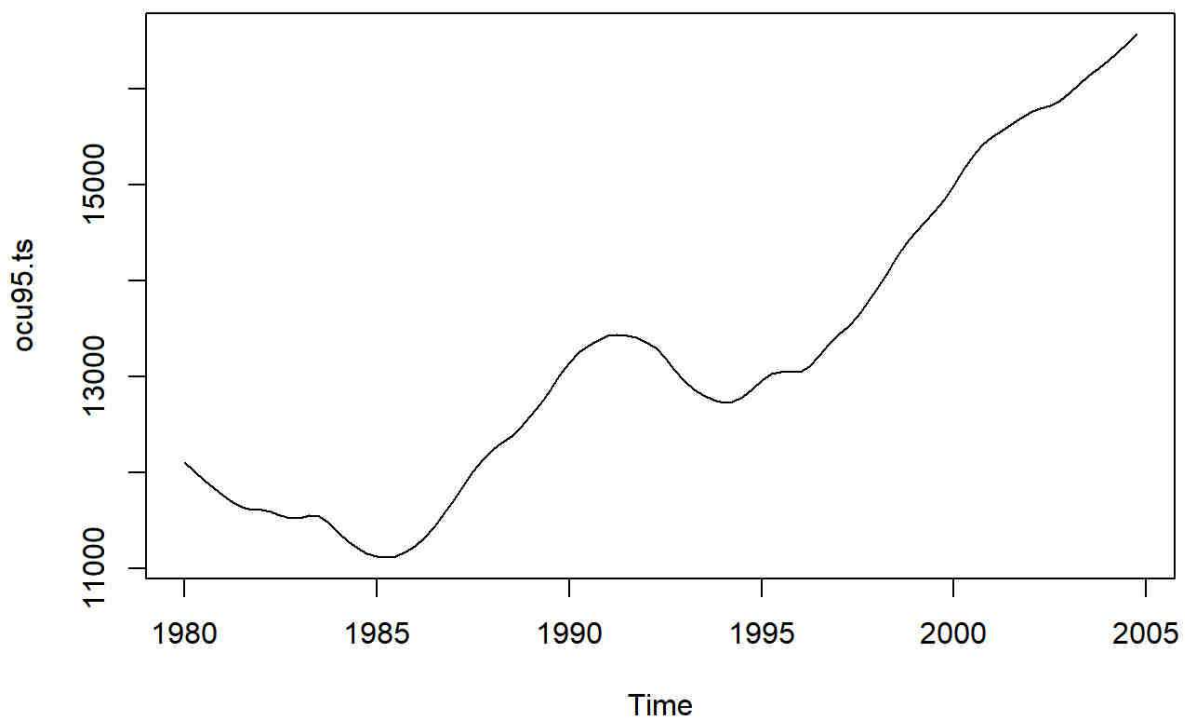
```

```
plot(value~tiempo,data = ocu,type="l")
```

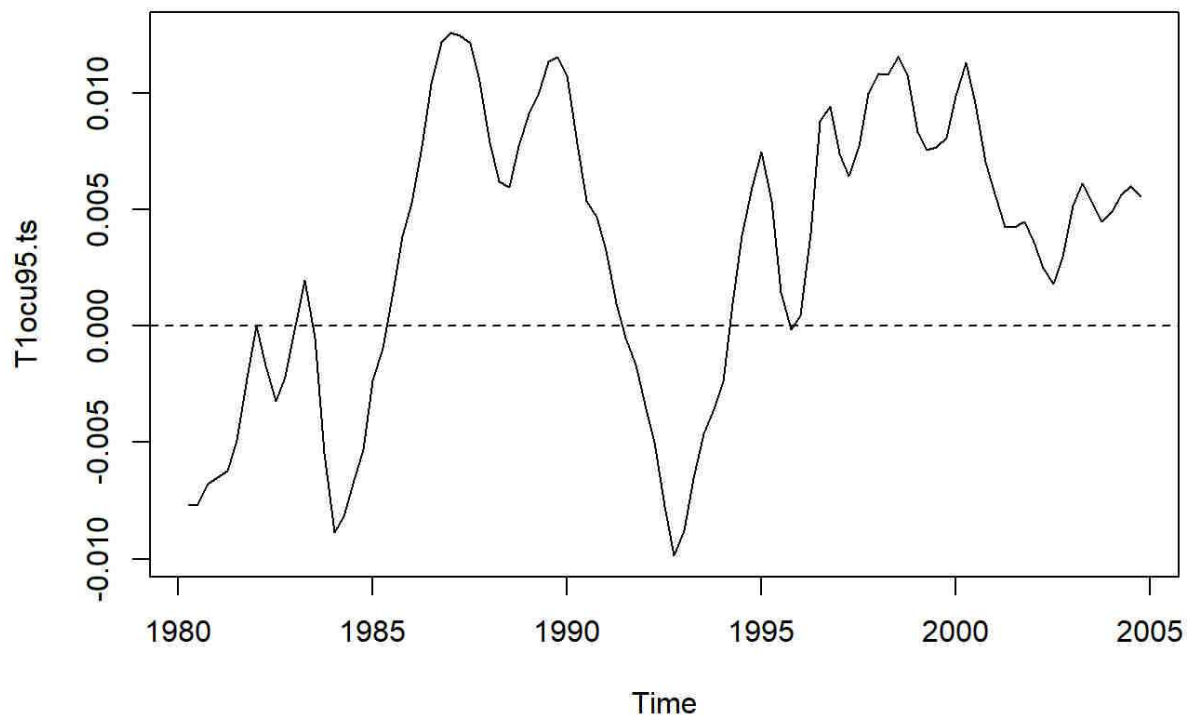


```
#Puedo convertirlo en objeto de serie temporal
ocu95.ts<-ts(ocu$value, start=c(1980,1), frequency = 4)

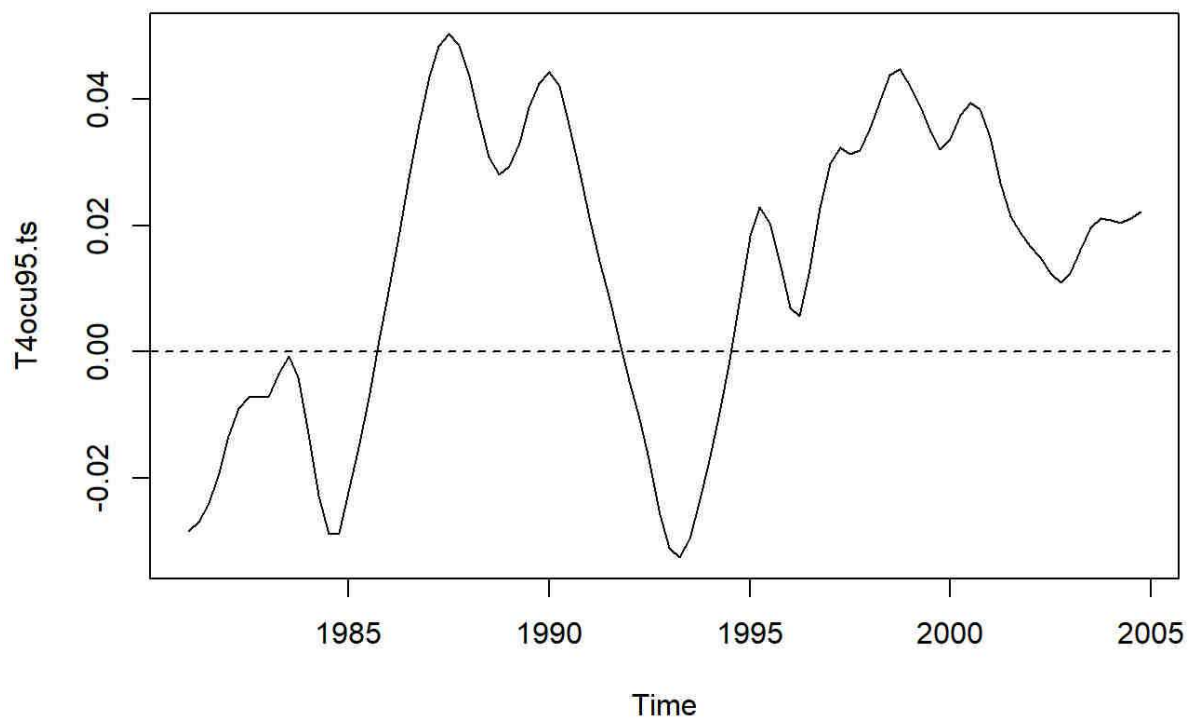
plot(ocu95.ts)
```



```
T1ocu95.ts<-ocu95.ts/stats::lag(ocu95.ts, k = -1)-1  
plot(T1ocu95.ts)  
abline(h=0,lty=2)
```



```
T4ocu95.ts<-ocu95.ts/stats::lag(ocu95.ts, k = -4)-1  
plot(T4ocu95.ts)  
abline(h=0,lty=2)
```

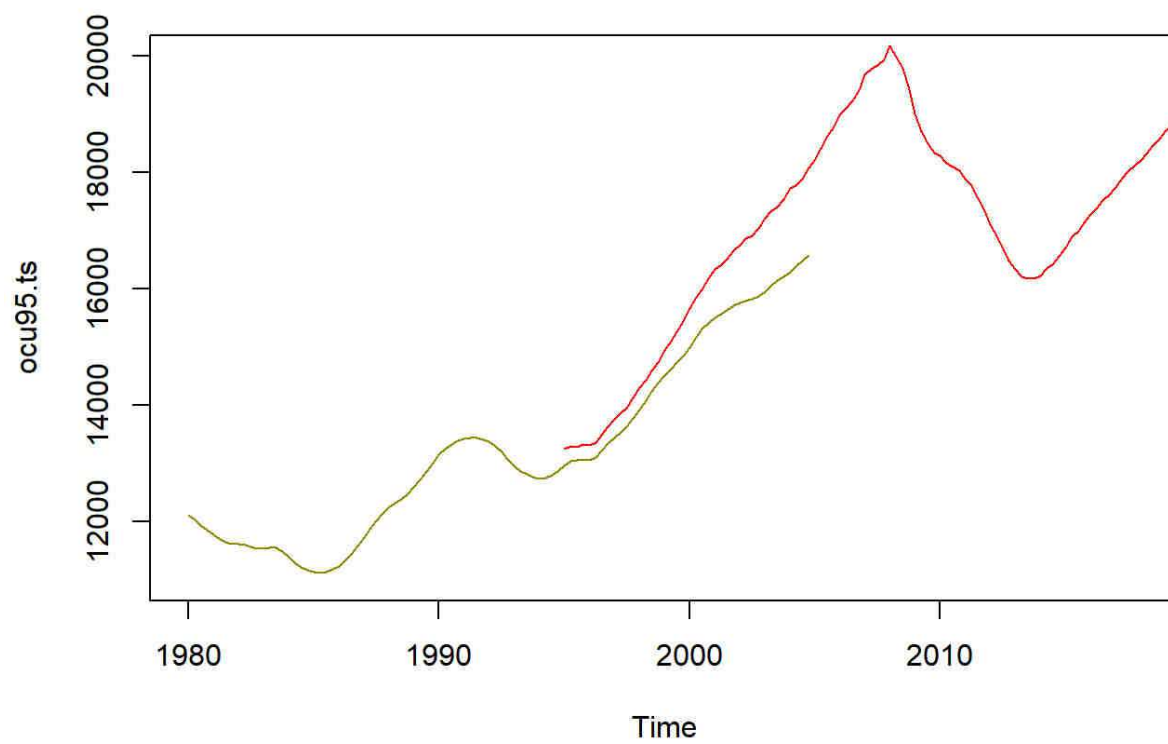


```
ts.union(ocu95.ts,T1ocu95.ts,T4ocu95.ts, dframe = TRUE)
```

##	ocu95.ts	T1ocu95.ts	T4ocu95.ts
## 1	12114.4	NA	NA
## 2	12021.1	-7.701578e-03	NA
## 3	11928.8	-7.678166e-03	NA
## 4	11848.0	-6.773523e-03	NA
## 5	11770.8	-6.515868e-03	-0.0283629400
## 6	11697.4	-6.235770e-03	-0.0269276522
## 7	11640.1	-4.898524e-03	-0.0242019315
## 8	11613.9	-2.250840e-03	-0.0197586090
## 9	11613.7	-1.722074e-05	-0.0133465865
## 10	11593.7	-1.722104e-03	-0.0088652179
## 11	11556.1	-3.243141e-03	-0.0072164328
## 12	11530.7	-2.197973e-03	-0.0071638295
## 13	11530.9	1.734500e-05	-0.0071295108
## 14	11553.4	1.951279e-03	-0.0034760258
## 15	11547.4	-5.193276e-04	-0.0007528491
## 16	11484.1	-5.481753e-03	-0.0040413852
## 17	11382.4	-8.855722e-03	-0.0128784397
## 18	11289.5	-8.161723e-03	-0.0228417609
## 19	11214.7	-6.625626e-03	-0.0288116806
## 20	11155.4	-5.287703e-03	-0.0286221820
## 21	11129.3	-2.339674e-03	-0.0222360838
## 22	11118.5	-9.704114e-04	-0.0151468178
## 23	11132.4	1.250169e-03	-0.0073385824
## 24	11175.1	3.835651e-03	0.0017659609
## 25	11234.1	5.279595e-03	0.0094165851
## 26	11320.7	7.708673e-03	0.0181859064
## 27	11438.3	1.038805e-02	0.0274783515
## 28	11577.4	1.216090e-02	0.0359996779
## 29	11722.8	1.255895e-02	0.0435014821
## 30	11868.8	1.245436e-02	0.0484157340
## 31	12012.8	1.213265e-02	0.0502259951
## 32	12138.9	1.049714e-02	0.0484996631
## 33	12234.5	7.875508e-03	0.0436499812
## 34	12310.4	6.203768e-03	0.0372067943
## 35	12383.8	5.962438e-03	0.0308837240
## 36	12479.4	7.719763e-03	0.0280503176
## 37	12592.9	9.094989e-03	0.0292942090
## 38	12717.4	9.886523e-03	0.0330614765
## 39	12861.6	1.133880e-02	0.0385826644
## 40	13010.0	1.153822e-02	0.0425180698
## 41	13149.7	1.073789e-02	0.0442153912
## 42	13252.8	7.840483e-03	0.0420998003
## 43	13323.8	5.357358e-03	0.0359364309
## 44	13386.1	4.675843e-03	0.0289085319
## 45	13429.2	3.219758e-03	0.0212552378
## 46	13442.1	9.605933e-04	0.0142837740
## 47	13435.3	-5.058733e-04	0.0083684835
## 48	13412.8	-1.674693e-03	0.0019946063
## 49	13367.4	-3.384826e-03	-0.0046019123
## 50	13300.6	-4.997232e-03	-0.0105266290
## 51	13197.6	-7.744012e-03	-0.0176921989
## 52	13067.4	-9.865430e-03	-0.0257515209
## 53	12952.6	-8.785221e-03	-0.0310307165
## 54	12867.9	-6.539228e-03	-0.0325323670
## 55	12808.3	-4.631680e-03	-0.0294977875
## 56	12761.5	-3.653881e-03	-0.0234094005
## 57	12730.8	-2.405673e-03	-0.0171239751
## 58	12742.8	9.425959e-04	-0.0097218660
## 59	12792.4	3.892394e-03	-0.0012413825
## 60	12868.4	5.941027e-03	0.0083767582

```
## 61 12964.5 7.467906e-03 0.0183570553
## 62 13033.8 5.345366e-03 0.0228364253
## 63 13052.3 1.419387e-03 0.0203167506
## 64 13050.0 -1.762142e-04 0.0141120885
## 65 13055.6 4.291188e-04 0.0070268811
## 66 13107.4 3.967646e-03 0.0056468566
## 67 13222.3 8.766041e-03 0.0130245244
## 68 13346.8 9.415911e-03 0.0227432950
## 69 13445.2 7.372554e-03 0.0298416005
## 70 13531.7 6.433523e-03 0.0323710271
## 71 13635.7 7.685657e-03 0.0312653623
## 72 13771.9 9.988486e-03 0.0318503312
## 73 13921.1 1.083365e-02 0.0353955315
## 74 14071.1 1.077501e-02 0.0398619538
## 75 14233.7 1.155560e-02 0.0438554676
## 76 14387.4 1.079832e-02 0.0446924535
## 77 14507.1 8.319780e-03 0.0420943747
## 78 14616.5 7.541135e-03 0.0387602959
## 79 14728.6 7.669415e-03 0.0347695961
## 80 14847.6 8.079519e-03 0.0319863214
## 81 14994.2 9.873650e-03 0.0335766625
## 82 15163.7 1.130437e-02 0.0374371430
## 83 15308.4 9.542526e-03 0.0393655880
## 84 15416.2 7.041886e-03 0.0382957515
## 85 15503.7 5.675847e-03 0.0339798055
## 86 15569.3 4.231248e-03 0.0267480892
## 87 15635.3 4.239112e-03 0.0213542891
## 88 15705.4 4.483445e-03 0.0187594868
## 89 15761.7 3.584754e-03 0.0166411889
## 90 15800.9 2.487041e-03 0.0148754279
## 91 15829.6 1.816352e-03 0.0124270081
## 92 15876.4 2.956487e-03 0.0108879748
## 93 15958.3 5.158600e-03 0.0124732738
## 94 16055.9 6.115940e-03 0.0161383212
## 95 16140.3 5.256635e-03 0.0196277859
## 96 16212.2 4.454688e-03 0.0211508906
## 97 16291.5 4.891378e-03 0.0208794170
## 98 16383.1 5.622564e-03 0.0203788016
## 99 16481.0 5.975670e-03 0.0211086535
## 100 16572.2 5.533645e-03 0.0222054996
```

```
plot(ocu95.ts, col="yellow4", xlim=c(1980,2018), ylim=c(11000,20000))
lines(ocu10.ts,type="l", col="red")
```



```
ts.union(ocu10.ts,ocu95.ts)
```

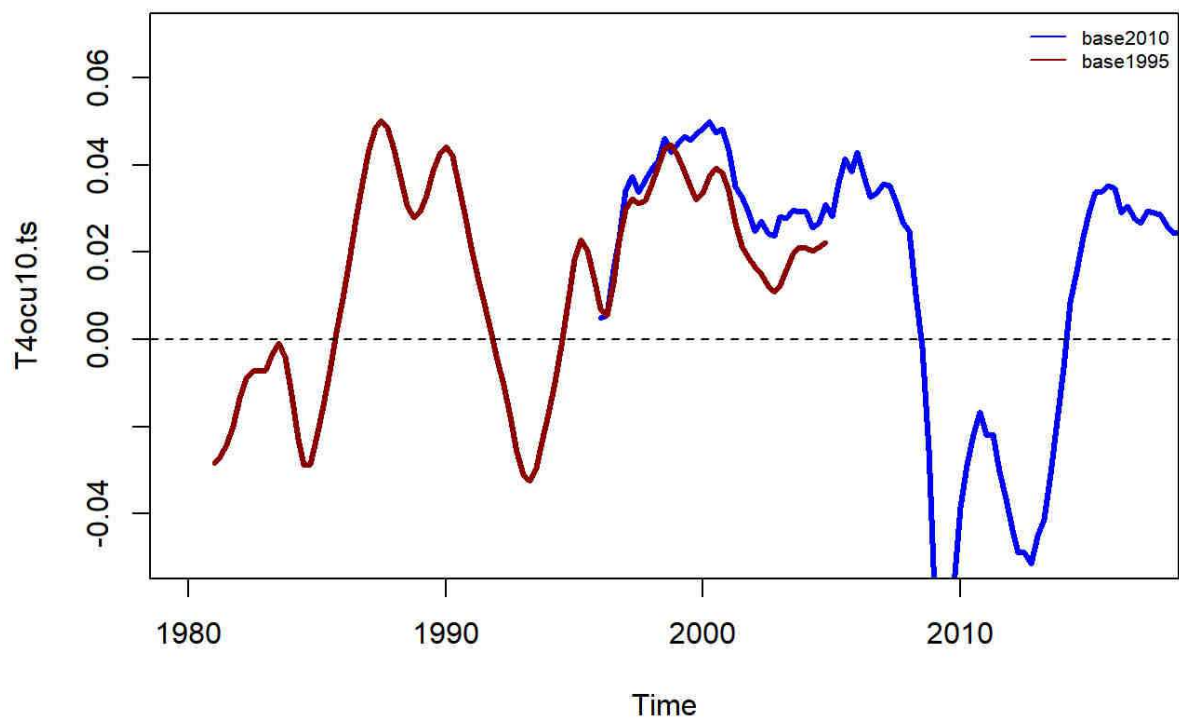


##		ocu10.ts	ocu95.ts
##	1980 Q1	NA	12114.4
##	1980 Q2	NA	12021.1
##	1980 Q3	NA	11928.8
##	1980 Q4	NA	11848.0
##	1981 Q1	NA	11770.8
##	1981 Q2	NA	11697.4
##	1981 Q3	NA	11640.1
##	1981 Q4	NA	11613.9
##	1982 Q1	NA	11613.7
##	1982 Q2	NA	11593.7
##	1982 Q3	NA	11556.1
##	1982 Q4	NA	11530.7
##	1983 Q1	NA	11530.9
##	1983 Q2	NA	11553.4
##	1983 Q3	NA	11547.4
##	1983 Q4	NA	11484.1
##	1984 Q1	NA	11382.4
##	1984 Q2	NA	11289.5
##	1984 Q3	NA	11214.7
##	1984 Q4	NA	11155.4
##	1985 Q1	NA	11129.3
##	1985 Q2	NA	11118.5
##	1985 Q3	NA	11132.4
##	1985 Q4	NA	11175.1
##	1986 Q1	NA	11234.1
##	1986 Q2	NA	11320.7
##	1986 Q3	NA	11438.3
##	1986 Q4	NA	11577.4
##	1987 Q1	NA	11722.8
##	1987 Q2	NA	11868.8
##	1987 Q3	NA	12012.8
##	1987 Q4	NA	12138.9
##	1988 Q1	NA	12234.5
##	1988 Q2	NA	12310.4
##	1988 Q3	NA	12383.8
##	1988 Q4	NA	12479.4
##	1989 Q1	NA	12592.9
##	1989 Q2	NA	12717.4
##	1989 Q3	NA	12861.6
##	1989 Q4	NA	13010.0
##	1990 Q1	NA	13149.7
##	1990 Q2	NA	13252.8
##	1990 Q3	NA	13323.8
##	1990 Q4	NA	13386.1
##	1991 Q1	NA	13429.2
##	1991 Q2	NA	13442.1
##	1991 Q3	NA	13435.3
##	1991 Q4	NA	13412.8
##	1992 Q1	NA	13367.4
##	1992 Q2	NA	13300.6
##	1992 Q3	NA	13197.6
##	1992 Q4	NA	13067.4
##	1993 Q1	NA	12952.6
##	1993 Q2	NA	12867.9
##	1993 Q3	NA	12808.3
##	1993 Q4	NA	12761.5
##	1994 Q1	NA	12730.8
##	1994 Q2	NA	12742.8
##	1994 Q3	NA	12792.4
##	1994 Q4	NA	12868.4

##	1995	Q1	13241.00	12964.5
##	1995	Q2	13292.30	13033.8
##	1995	Q3	13282.90	13052.3
##	1995	Q4	13323.70	13050.0
##	1996	Q1	13305.30	13055.6
##	1996	Q2	13366.00	13107.4
##	1996	Q3	13496.90	13222.3
##	1996	Q4	13632.50	13346.8
##	1997	Q1	13760.30	13445.2
##	1997	Q2	13866.40	13531.7
##	1997	Q3	13952.60	13635.7
##	1997	Q4	14131.50	13771.9
##	1998	Q1	14298.00	13921.1
##	1998	Q2	14435.40	14071.1
##	1998	Q3	14595.80	14233.7
##	1998	Q4	14739.50	14387.4
##	1999	Q1	14940.70	14507.1
##	1999	Q2	15106.80	14616.5
##	1999	Q3	15264.80	14728.6
##	1999	Q4	15434.40	14847.6
##	2000	Q1	15663.40	14994.2
##	2000	Q2	15860.90	15163.7
##	2000	Q3	15990.10	15308.4
##	2000	Q4	16180.10	15416.2
##	2001	Q1	16343.60	15503.7
##	2001	Q2	16417.10	15569.3
##	2001	Q3	16512.40	15635.3
##	2001	Q4	16653.40	15705.4
##	2002	Q1	16748.10	15761.7
##	2002	Q2	16862.60	15800.9
##	2002	Q3	16915.40	15829.6
##	2002	Q4	17048.40	15876.4
##	2003	Q1	17220.70	15958.3
##	2003	Q2	17332.40	16055.9
##	2003	Q3	17417.90	16140.3
##	2003	Q4	17546.00	16212.2
##	2004	Q1	17726.00	16291.5
##	2004	Q2	17777.90	16383.1
##	2004	Q3	17884.30	16481.0
##	2004	Q4	18089.20	16572.2
##	2005	Q1	18226.10	NA
##	2005	Q2	18415.40	NA
##	2005	Q3	18627.40	NA
##	2005	Q4	18784.70	NA
##	2006	Q1	19009.20	NA
##	2006	Q2	19109.40	NA
##	2006	Q3	19238.30	NA
##	2006	Q4	19413.50	NA
##	2007	Q1	19686.30	NA
##	2007	Q2	19783.30	NA
##	2007	Q3	19841.60	NA
##	2007	Q4	19938.50	NA
##	2008	Q1	20175.10	NA
##	2008	Q2	19988.70	NA
##	2008	Q3	19803.90	NA
##	2008	Q4	19431.60	NA
##	2009	Q1	19014.90	NA
##	2009	Q2	18711.90	NA
##	2009	Q3	18503.40	NA
##	2009	Q4	18338.20	NA
##	2010	Q1	18286.10	NA
##	2010	Q2	18170.80	NA

##	2010	Q3	18101.00	NA
##	2010	Q4	18034.40	NA
##	2011	Q1	17885.40	NA
##	2011	Q2	17775.00	NA
##	2011	Q3	17553.40	NA
##	2011	Q4	17374.90	NA
##	2012	Q1	17103.71	NA
##	2012	Q2	16907.10	NA
##	2012	Q3	16696.96	NA
##	2012	Q4	16481.13	NA
##	2013	Q1	16333.24	NA
##	2013	Q2	16210.47	NA
##	2013	Q3	16184.64	NA
##	2013	Q4	16174.91	NA
##	2014	Q1	16229.45	NA
##	2014	Q2	16349.74	NA
##	2014	Q3	16439.06	NA
##	2014	Q4	16555.60	NA
##	2015	Q1	16710.73	NA
##	2015	Q2	16902.33	NA
##	2015	Q3	16998.61	NA
##	2015	Q4	17138.29	NA
##	2016	Q1	17290.98	NA
##	2016	Q2	17394.23	NA
##	2016	Q3	17516.82	NA
##	2016	Q4	17612.97	NA
##	2017	Q1	17752.95	NA
##	2017	Q2	17906.29	NA
##	2017	Q3	18027.35	NA
##	2017	Q4	18118.42	NA
##	2018	Q1	18215.95	NA
##	2018	Q2	18346.26	NA
##	2018	Q3	18467.90	NA
##	2018	Q4	18587.88	NA
##	2019	Q1	18726.17	NA
##	2019	Q2	18805.48	NA

```
plot(T4ocu10.ts, col="blue", xlim=c(1980,2017), ylim=c(-0.05,0.07),lwd=3)
lines(T4ocu95.ts, col="darkred",lwd=3)
abline(h=0,lty=2)
legend("topright",legend = c("base2010", "base1995"),lty=c(1), cex=0.65, lwd=c(1,1),
      bty="n",
      col=c("blue", "darkred"))
```



```
ts.union(T4ocu95.ts,T4ocu10.ts)
```

##		T4ocu95.ts	T4ocu10.ts
## 1981	Q1	-0.0283629400	NA
## 1981	Q2	-0.0269276522	NA
## 1981	Q3	-0.0242019315	NA
## 1981	Q4	-0.0197586090	NA
## 1982	Q1	-0.0133465865	NA
## 1982	Q2	-0.0088652179	NA
## 1982	Q3	-0.0072164328	NA
## 1982	Q4	-0.0071638295	NA
## 1983	Q1	-0.0071295108	NA
## 1983	Q2	-0.0034760258	NA
## 1983	Q3	-0.0007528491	NA
## 1983	Q4	-0.0040413852	NA
## 1984	Q1	-0.0128784397	NA
## 1984	Q2	-0.0228417609	NA
## 1984	Q3	-0.0288116806	NA
## 1984	Q4	-0.0286221820	NA
## 1985	Q1	-0.0222360838	NA
## 1985	Q2	-0.0151468178	NA
## 1985	Q3	-0.0073385824	NA
## 1985	Q4	0.0017659609	NA
## 1986	Q1	0.0094165851	NA
## 1986	Q2	0.0181859064	NA
## 1986	Q3	0.0274783515	NA
## 1986	Q4	0.0359996779	NA
## 1987	Q1	0.0435014821	NA
## 1987	Q2	0.0484157340	NA
## 1987	Q3	0.0502259951	NA
## 1987	Q4	0.0484996631	NA
## 1988	Q1	0.0436499812	NA
## 1988	Q2	0.0372067943	NA
## 1988	Q3	0.0308837240	NA
## 1988	Q4	0.0280503176	NA
## 1989	Q1	0.0292942090	NA
## 1989	Q2	0.0330614765	NA
## 1989	Q3	0.0385826644	NA
## 1989	Q4	0.0425180698	NA
## 1990	Q1	0.0442153912	NA
## 1990	Q2	0.0420998003	NA
## 1990	Q3	0.0359364309	NA
## 1990	Q4	0.0289085319	NA
## 1991	Q1	0.0212552378	NA
## 1991	Q2	0.0142837740	NA
## 1991	Q3	0.0083684835	NA
## 1991	Q4	0.0019946063	NA
## 1992	Q1	-0.0046019123	NA
## 1992	Q2	-0.0105266290	NA
## 1992	Q3	-0.0176921989	NA
## 1992	Q4	-0.0257515209	NA
## 1993	Q1	-0.0310307165	NA
## 1993	Q2	-0.0325323670	NA
## 1993	Q3	-0.0294977875	NA
## 1993	Q4	-0.0234094005	NA
## 1994	Q1	-0.0171239751	NA
## 1994	Q2	-0.0097218660	NA
## 1994	Q3	-0.0012413825	NA
## 1994	Q4	0.0083767582	NA
## 1995	Q1	0.0183570553	NA
## 1995	Q2	0.0228364253	NA
## 1995	Q3	0.0203167506	NA
## 1995	Q4	0.0141120885	NA

##	1996	Q1	0.0070268811	0.004856129
##	1996	Q2	0.0056468566	0.005544563
##	1996	Q3	0.0130245244	0.016110940
##	1996	Q4	0.0227432950	0.023176745
##	1997	Q1	0.0298416005	0.034196899
##	1997	Q2	0.0323710271	0.037438276
##	1997	Q3	0.0312653623	0.033763309
##	1997	Q4	0.0318503312	0.036603704
##	1998	Q1	0.0353955315	0.039076183
##	1998	Q2	0.0398619538	0.041034443
##	1998	Q3	0.0438554676	0.046098935
##	1998	Q4	0.0446924535	0.043024449
##	1999	Q1	0.0420943747	0.044950343
##	1999	Q2	0.0387602959	0.046510661
##	1999	Q3	0.0347695961	0.045835103
##	1999	Q4	0.0319863214	0.047145426
##	2000	Q1	0.0335766625	0.048371228
##	2000	Q2	0.0374371430	0.049917918
##	2000	Q3	0.0393655880	0.047514543
##	2000	Q4	0.0382957515	0.048314155
##	2001	Q1	0.0339798055	0.043426076
##	2001	Q2	0.0267480892	0.035067367
##	2001	Q3	0.0213542891	0.032663961
##	2001	Q4	0.0187594868	0.029251982
##	2002	Q1	0.0166411889	0.024749749
##	2002	Q2	0.0148754279	0.027136340
##	2002	Q3	0.0124270081	0.024405901
##	2002	Q4	0.0108879748	0.023718880
##	2003	Q1	0.0124732738	0.028218126
##	2003	Q2	0.0161383212	0.027860472
##	2003	Q3	0.0196277859	0.029706658
##	2003	Q4	0.0211508906	0.029187490
##	2004	Q1	0.0208794170	0.029342594
##	2004	Q2	0.0203788016	0.025703307
##	2004	Q3	0.0211086535	0.026777051
##	2004	Q4	0.0222054996	0.030958623
##	2005	Q1	NA	0.028212795
##	2005	Q2	NA	0.035859128
##	2005	Q3	NA	0.041550410
##	2005	Q4	NA	0.038448356
##	2006	Q1	NA	0.042965857
##	2006	Q2	NA	0.037685850
##	2006	Q3	NA	0.032795774
##	2006	Q4	NA	0.033474051
##	2007	Q1	NA	0.035619595
##	2007	Q2	NA	0.035265367
##	2007	Q3	NA	0.031359320
##	2007	Q4	NA	0.027043037
##	2008	Q1	NA	0.024829450
##	2008	Q2	NA	0.010382494
##	2008	Q3	NA	-0.001900048
##	2008	Q4	NA	-0.025423176
##	2009	Q1	NA	-0.057506530
##	2009	Q2	NA	-0.063876090
##	2009	Q3	NA	-0.065668883
##	2009	Q4	NA	-0.056269170
##	2010	Q1	NA	-0.038327838
##	2010	Q2	NA	-0.028917427
##	2010	Q3	NA	-0.021747355
##	2010	Q4	NA	-0.016566511
##	2011	Q1	NA	-0.021912819
##	2011	Q2	NA	-0.021782200

```
## 2011 Q3      NA -0.030252472
## 2011 Q4      NA -0.036569001
## 2012 Q1      NA -0.043705425
## 2012 Q2      NA -0.048827286
## 2012 Q3      NA -0.048790491
## 2012 Q4      NA -0.051440411
## 2013 Q1      NA -0.045047125
## 2013 Q2      NA -0.041203235
## 2013 Q3      NA -0.030683787
## 2013 Q4      NA -0.018580039
## 2014 Q1      NA -0.006354343
## 2014 Q2      NA  0.008591239
## 2014 Q3      NA  0.015720219
## 2014 Q4      NA  0.023536270
## 2015 Q1      NA  0.029654669
## 2015 Q2      NA  0.033798466
## 2015 Q3      NA  0.034037711
## 2015 Q4      NA  0.035195392
## 2016 Q1      NA  0.034723019
## 2016 Q2      NA  0.029102197
## 2016 Q3      NA  0.030485730
## 2016 Q4      NA  0.027697519
## 2017 Q1      NA  0.026717519
## 2017 Q2      NA  0.029438445
## 2017 Q3      NA  0.029145009
## 2017 Q4      NA  0.028697368
## 2018 Q1      NA  0.026080456
## 2018 Q2      NA  0.024571036
## 2018 Q3      NA  0.024437976
## 2018 Q4      NA  0.025910703
## 2019 Q1      NA  0.028009348
## 2019 Q2      NA  0.025030439
```

## 5.3 Reconstruyo la serie de ocupados por Tasas de crecimiento

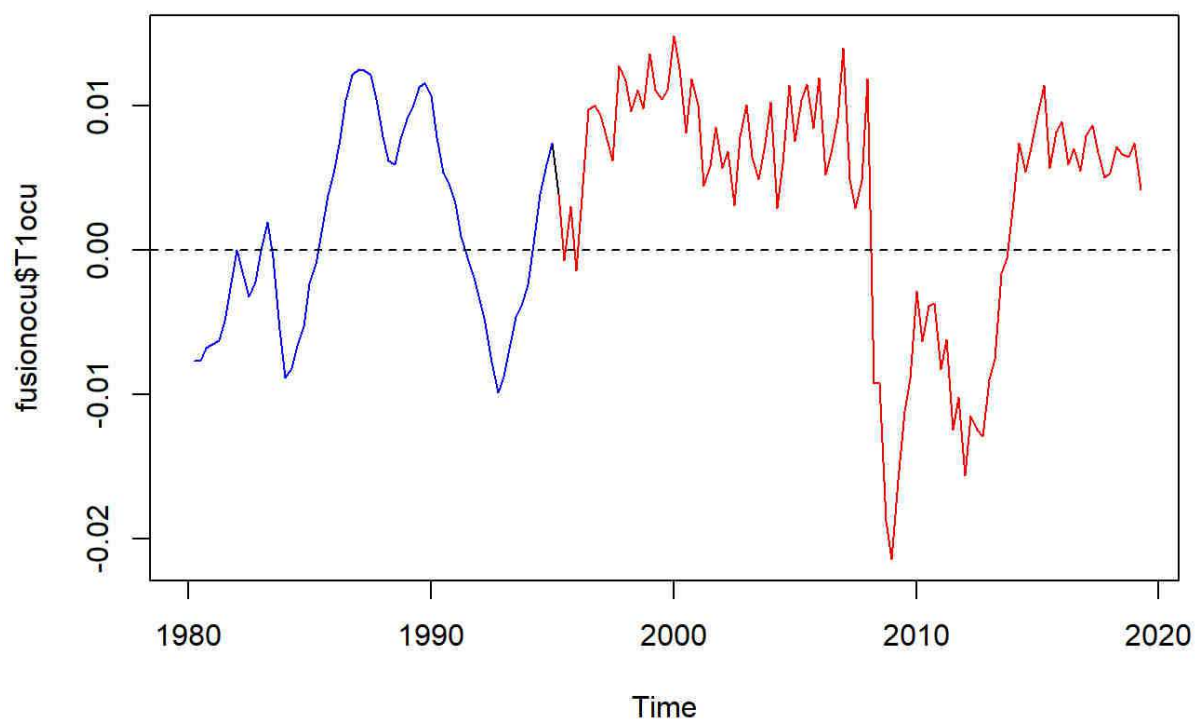
```
# Reconstruyo la serie de ocupado por tasa de crecimiento ----

#creo la variable tiempo
tiempo<-ts(seq(1980,tiempo_max,by=0.25), start = c(1980,1),frequency = 4)

fusionocu<-ts.union(tiempo,T1ocu95.ts,T1ocu10.ts,ocu10.ts, dframe = TRUE)

fusionocu$T1ocu<-fusionocu$T1ocu10.ts
fusionocu$T1ocu[tiempo>1980 &tiempo<=1995]<-fusionocu$T1ocu95.ts[tiempo>1980 &tiempo<=1995]

plot(fusionocu$T1ocu)
lines(fusionocu$T1ocu10.ts, col="red")
lines(window(fusionocu$T1ocu95.ts,start=1980, end=1995), col="blue")
abline(h=0,lty=2)
```

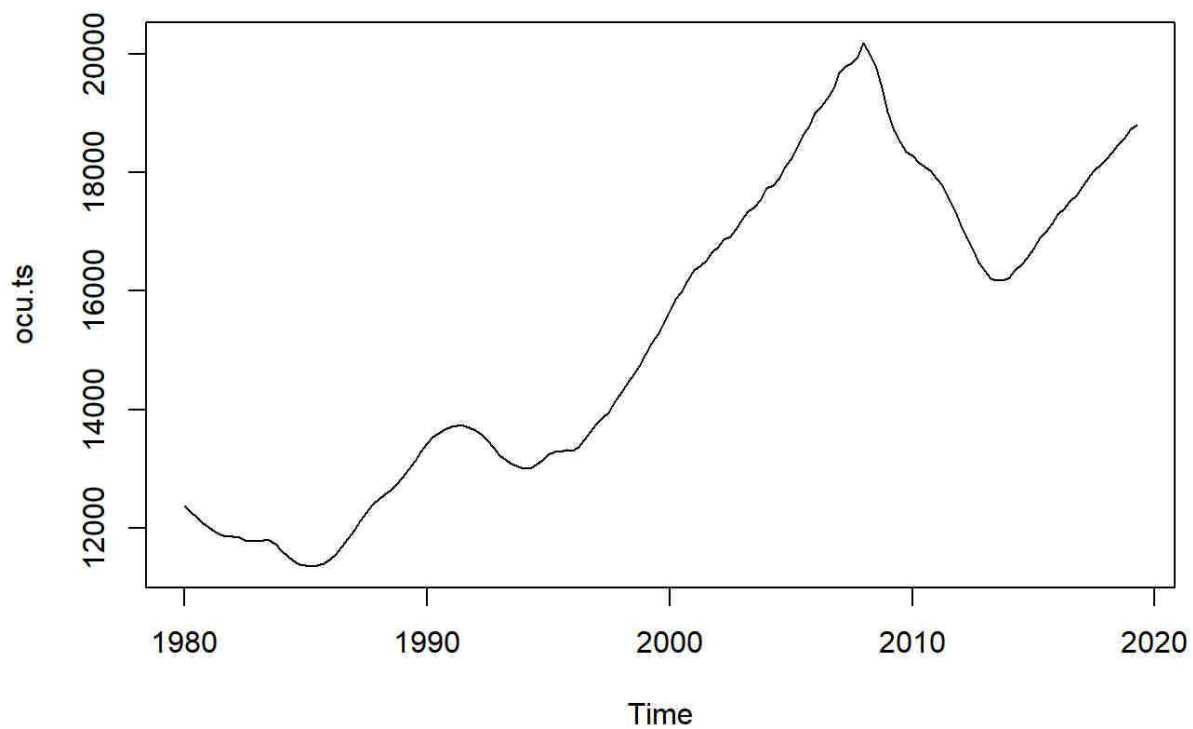


```
# Reconstruyo la serie de OCUPADOS

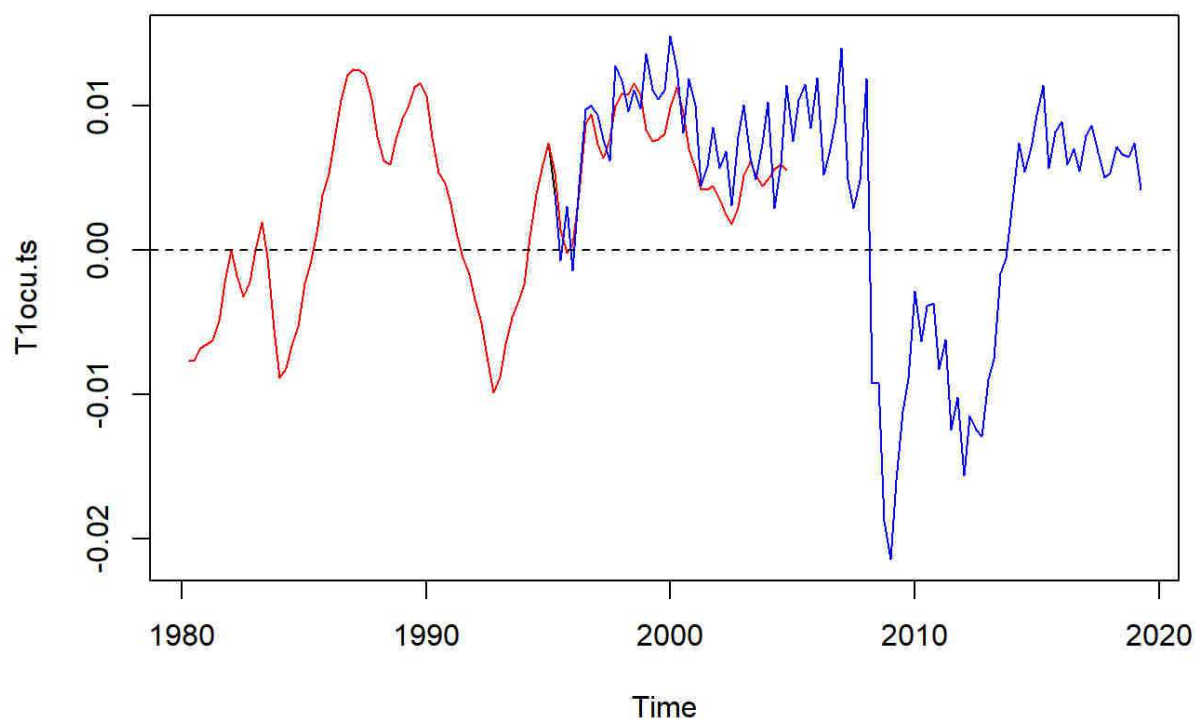
fusionocu$ocu<-fusionocu$ocu10.ts
for(i in seq(1994.75, 1980, by=-0.25)){
  fusionocu$ocu[tiempo==i]<-fusionocu$ocu[tiempo==(i+0.25)]/(1+fusionocu$T1ocu[tiempo==(i+0.25)])
}

ocu.ts<-fusionocu$ocu
plot(ocu.ts)
```





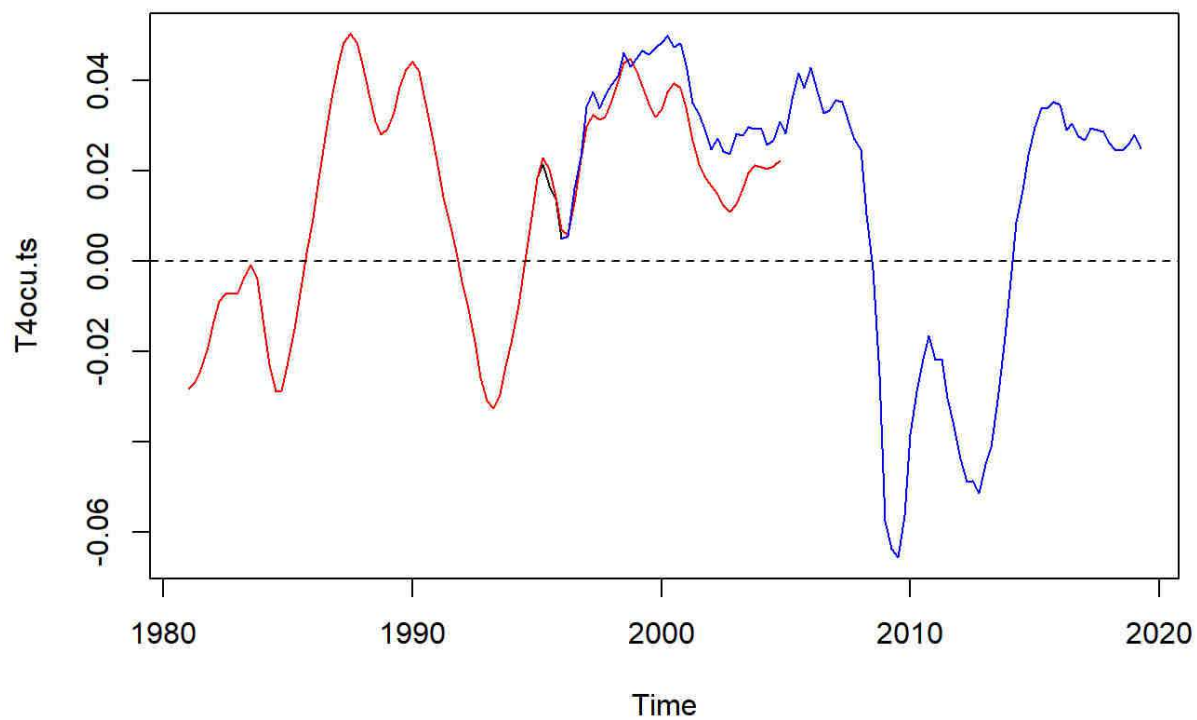
```
T1ocu.ts<-ocu.ts/stats::lag(ocu.ts, k = -1)-1  
plot(T1ocu.ts)  
lines(T1ocu95.ts,col="red")  
lines(T1ocu10.ts,col="blue")  
  
abline(h=0,lty=2)
```



```

T4ocu.ts<-ocu.ts/stats::lag(ocu.ts, k = -4)-1
plot(T4ocu.ts)
lines(T4ocu95.ts,col="red")
lines(T4ocu10.ts,col="blue")
abline(h=0,lty=2)

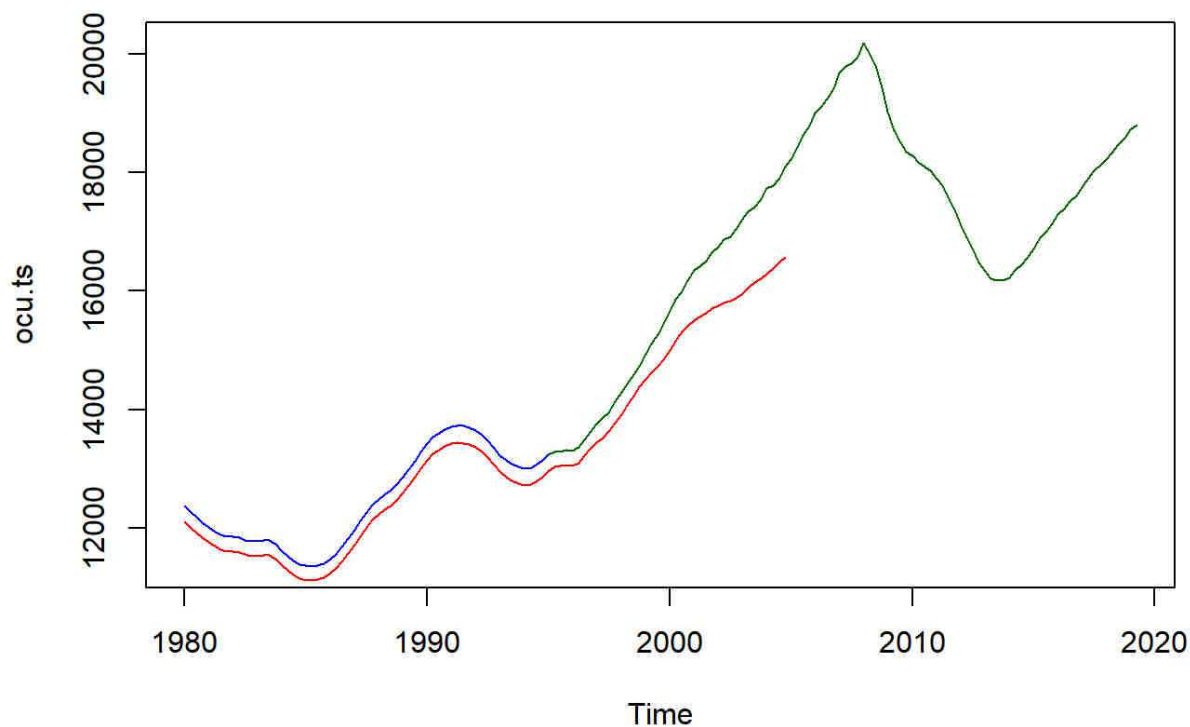
```



```

plot(ocu.ts, col="blue")
lines(ocu10.ts, col="darkgreen")
lines(ocu95.ts, col="red")

```



## 6 Fusiono las dos series largas en un dataframe

```
# FUSIONO LOS DATOS FINALES

okun<-ts.union(tiempo=time(pib.ts),pib.ts,ocu.ts,"T1pib.ts"=T1.ts,"T4pib.ts"=T4.ts,
               T1ocu.ts,T4ocu.ts, dframe = TRUE)

#View(okun)

# guardo los datos

save(okun, file="okundat.Rdata")
```

Con esto termino esta práctica 1, en la que el objetivo principal era bajarnos series de diferentes bases del PIB y de ocupados y enlazarlas utilizando tasas de crecimiento para construir dos series largas, desde 1970 y 1980 hasta el último dato publicado para las series de PIB y Ocupados (según datos de Contabilidad Nacional)



Proyecto de innovación  
“Innova Docencia” de la UCM  
Convocatoria 2018-2019  
Proyecto número 275

**R-adaptación de la asignatura de métodos econométricos en economía y finanzas del  
Grado de Estadística Aplicada**

Facultad de Estudios Estadísticos  
Departamento de Economía Aplicada, Pública y Política

**Título**

***Práctica 2 : La Ley de Okun: ¿a qué ritmo debe crecer la Economía para  
comenzar a generar empleo***

**Correspondiente al**

**Capítulo 2 REGRESIÓN LINEAL**

- El Modelo Lineal General (MLG)
- Hipótesis del modelo
- Estimación por Mínimos Cuadrados Ordinarios (MCO)
- Inferencia y predicción en el MLG
- Contrastes de cambio estructural
- Multicolinealidad y errores de medida

***Práctica 2 : La Ley de Okun: ¿a qué ritmo debe crecer la Economía para  
comenzar a generar empleo***

***Práctica 3 : La Ley de Okun: Test de Cambio Estructural, ¿han sido  
“eficaces” las reformas del mercado de trabajo?***

***Práctica 4 : Ley de Okun con datos de la UE desde la web de Eurostat***

***Práctica 5 : La ecuación de salarios y la discriminación por razón de género***

# P2- Ley de Okun: tasa de crecimiento de los ocupados y tasa de crecimiento del PIB

## Asignatura Métodos Econométricos en Economía y Finanzas

Profesores  
Loren Escot, Alicia Pérez-Alonso y Julio E. Sandubete

Dpto Economía Aplicada, Pública y Política  
Facultad Estudios Estadísticos - Universidad Complutense de Madrid

Curso académico 2018/2019

### Abstract

En esta práctica se muestra cómo aplicar una regresión en R para analizar la relación entre la tasa de crecimiento de los ocupados y del PIB (Ley de Okun).

Se utilizará esta Ley de Okun para proponer la primera práctica a entregar por los alumnos en su evaluación continua

- 1 Enunciado de la Tarea 1: Ley de Okun (primera parte)
- 2 Librerías necesarias
- 3 Carga de datos
- 4 Correlaciones
- 5 Estimación del Modelo Lineal
- 6 Diagnóstico de los residuos
- 7 Tarea Punto 2: ¿A qué ritmo debe crecer la economía para empezar a generar empleo?
- 8 Continuamos con la diagnosis del modelo
  - 8.1 Test de Normalidad sobre los residuos
  - 8.2 Contraste de Hipótesis
  - 8.3 Contraste de Especificación de Ramsey
  - 8.4 Contraste de hipótesis lineales sobre los coeficientes
  - 8.5 Contraste sobre un único parámetro
  - 8.6 Contraste de hipótesis por Sumas Residuales
- 9 Predicción en el modelo lineal
- 10 Test de Cambio Estructural
  - 10.1 Test de Chow de Cambio Estructural
  - 10.2 Test de Cambio Estructural mediante mínimos Cuadrados Recursivos

## 1 Enunciado de la Tarea 1: Ley de Okun (primera parte)

Tenéis que dar respuesta razonada a las siguientes tres preguntas:

1. ¿Hemos salido de la crisis económica? ¿cuándo?
2. ¿A qué ritmo tiene que crecer la economía española para comenzar a generar empleo?
3. ¿Se van a crear 2 millones de empleos durante el 2019 y 2020 en la economía Española?, ¿no?, ¿cuantos entonces?
4. ¿Han sido eficaces las reformas del mercado de trabajo? \* 1980 Estatuto de los trabajadores; \* 1984: inicio de la temporalidad; \* 1988: Nuevos contratos Temporales y basura; \* 1992: reforma del seguro desempleo; \* 1994: reforma importante que da entrada entre otras a la ETT; \* 1997: reforma pactada; \* 2001-2002: reforma Aznar

con huelga y marcha atrás; \* 2006: reforma leve; 2009: reforma leve; \* 2010 reforma importante ante la crisis continuada en 2012, con una Reforma importante

El trabajo se presentará en Word y contará al menos con los siguientes apartados:

- Introducción (¿qué se va a hacer?)
- Metodología y datos (¿cómo se va a hacer?: fundamentación en la Ley de Okun, datos, métodos de regresión por MCO. ¿qué datos se van a utilizar?)
- Resultados y diagnóstico
- Conclusiones al trabajo

## 2 Librerías necesarias

Inicialmente sólo voy a cargar la librería `stargazer`. Posteriormente cuando se revisen los diferentes test se hará referencia a más librerías

```
library(stargazer) # para hacer tablas
```

```
##  
## Please cite as:
```

```
## Hlavac, Marek (2018). stargazer: Well-Formatted Regression and Summary Statistics Tables.
```

```
## R package version 5.2.2. https://CRAN.R-project.org/package=stargazer
```

## 3 Carga de datos

Estos datos corresponden a las series largas de PIB y empleo generadas desde la práctica anterior

```
load(file="okundat.Rdata")
```

```
head(okun)
```

```
##      tiempo  pib.ts  ocu.ts    T1pib.ts    T4pib.ts  T1ocu.ts  T4ocu.ts  
## 1 1970.00 32.18634    NA          NA          NA      NA      NA  
## 2 1970.25 32.23748    NA 0.001588640          NA      NA      NA  
## 3 1970.50 32.40224    NA 0.005110833          NA      NA      NA  
## 4 1970.75 32.65790    NA 0.007890276          NA      NA      NA  
## 5 1971.00 32.91293    NA 0.007809178 0.02257438      NA      NA  
## 6 1971.25 33.55493    NA 0.019505927 0.04086708      NA      NA
```

```
tail(okun)
```

```
##      tiempo  pib.ts  ocu.ts   T1pib.ts  T4pib.ts   T1ocu.ts
## 193 2018.00 107.2630 18215.95 0.006236503 0.02894511 0.005383140
## 194 2018.25 107.9003 18346.26 0.005941471 0.02608961 0.007153510
## 195 2018.50 108.4781 18467.90 0.005354943 0.02503189 0.006630342
## 196 2018.75 109.0791 18587.88 0.005540289 0.02327338 0.006496514
## 197 2019.00 109.8059 18726.17 0.006663055 0.02370715 0.007439794
## 198 2019.25 110.3294 18805.48 0.004767503 0.02251245 0.004235036
##      T4ocu.ts
## 193 0.02608046
## 194 0.02457104
## 195 0.02443798
## 196 0.02591070
## 197 0.02800935
## 198 0.02503044
```

Consulto el último dato disponible. A media que vaya pasando las semanas irán apareciendo nuevos datos oficiales de PIB y Empleo y os pediré que compareis las predicciones que hacemos ahora con los datos oficiales que vayan saliendo.

```
# último dato disponible
tiempo_max<-max(okun$tiempo)
print(tiempo_max)
```

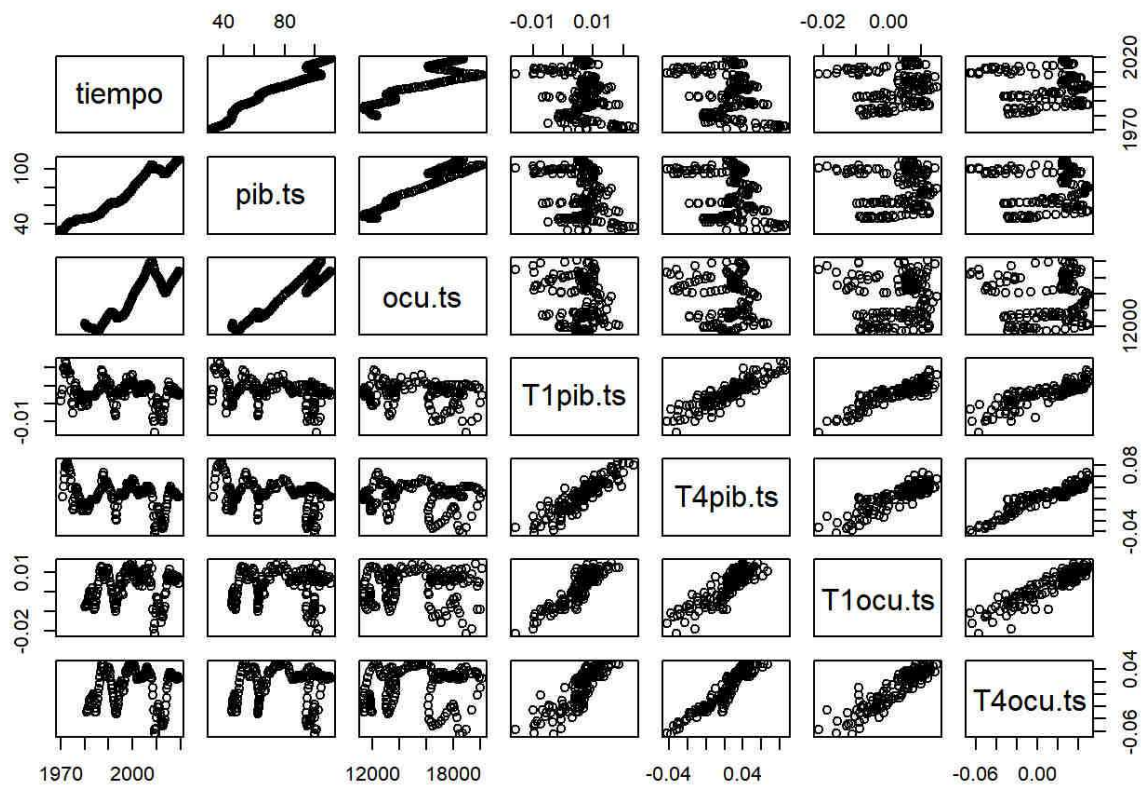
```
## [1] 2019.25
```

## 4 Correlaciones

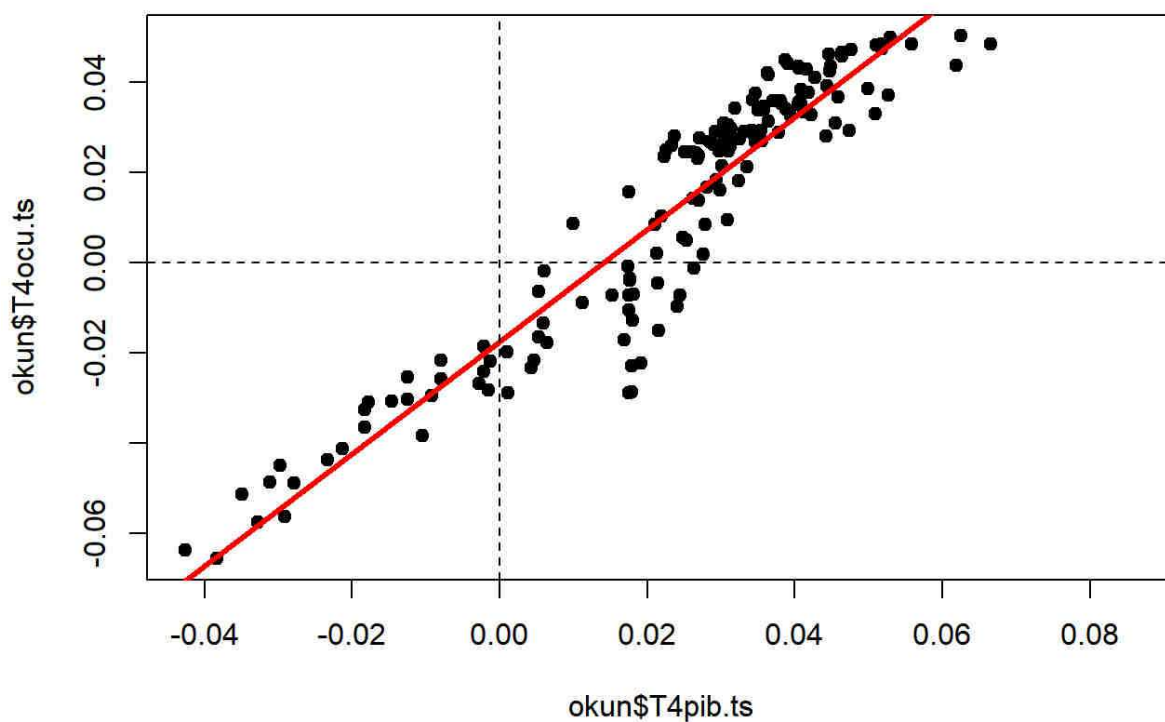
```
#correlaciones ----
cor.test(okun$T4pib.ts, okun$T4ocu.ts)
```

```
##
## Pearson's product-moment correlation
##
## data: okun$T4pib.ts and okun$T4ocu.ts
## t = 34.055, df = 152, p-value < 2.2e-16
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
## 0.9187478 0.9562319
## sample estimates:
## cor
## 0.9402771
```

```
pairs(okun)
```



```
plot(okun$T4pib.ts, okun$T4ocu.ts, pch=19)
abline(lm(T4ocu.ts ~ T4pib.ts, data=okun), col=2, lwd=3)
abline(h=0, lty=2, col="grey2")
abline(v=0, lty=2, col="grey2")
```





```
# para poder trabajar con el modelo de regresión cuando trabajamos con series temporales
#conviene asegurarse de que todas las series tienen el mismo rango temporal
```

```
T4pib.ts<-window(okun[,5], start=1981, end=tiempo_max)
T4ocu.ts<-window(okun[,7], start=1981, end=tiempo_max)
```

```
datokun<-ts.intersect(T4pib.ts,T4ocu.ts,dframe = TRUE)
```

## 5 Estimación del Modelo Lineal

```
# Estimación del Modelo Lineal ----
# Especificaciones del modelo I(A^2), A:B para interacciones, a*b, ~. , 0 o -1 para no poner término
cte
modelo<-T4ocu.ts ~ T4pib.ts
```

```
modelo.lm<-lm(modelo, data=datokun, na.action=NULL) #ojo, na.action=NULL para las series
salida.lm<-summary(modelo.lm)
print(salida.lm)
```

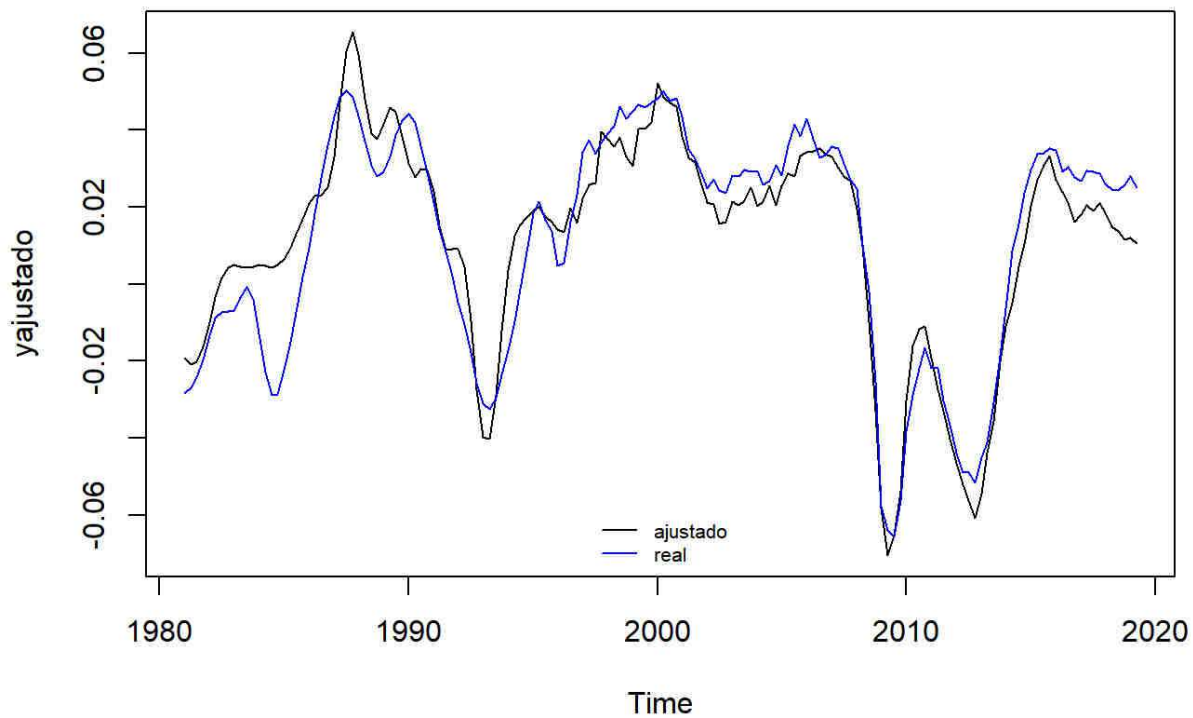
```
##
## Call:
## lm(formula = modelo, data = datokun, na.action = NULL)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.033511 -0.005997  0.002390  0.007533  0.015968
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -0.017443   0.001176  -14.84  <2e-16 ***
## T4pib.ts     1.243711   0.036521   34.05  <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.0101 on 152 degrees of freedom
## Multiple R-squared:  0.8841, Adjusted R-squared:  0.8834
## F-statistic: 1160 on 1 and 152 DF, p-value: < 2.2e-16
```

```
stargazer(modelo.lm,type="text")
```

```
##
## =====
##                      Dependent variable:
##                      -----
##                      T4ocu.ts
## -----
## T4pib.ts                1.244***
##                        (0.037)
##
## Constant                -0.017***
##                        (0.001)
## -----
## Observations              154
## R2                      0.884
## Adjusted R2              0.883
## Residual Std. Error      0.010 (df = 152)
## F Statistic              1,159.712*** (df = 1; 152)
## =====
## Note:                    *p<0.1; **p<0.05; ***p<0.01
```

```
SSR<- deviance(modelo.lm)
sigma_err<-salida.lm$sigma
R2<-salida.lm$r.squared
LogVero<-logLik(modelo.lm)
grLibtd<-modelo.lm$df
Anova_lm<-anova(modelo.lm)
AIC_lm<-AIC(modelo.lm)
BIC_lm<-AIC(modelo.lm, k=log(nrow(datokun)))
yajustado<-fitted(modelo.lm)

plot(yajustado)
lines(T4ocu.ts,col="blue")
legend("bottom",legend = c("ajustado", "real"),lty=c(1), cex=0.65, lwd=c(1,1),
      bty="n",
      col=c("black","blue"))
```



## 6 Diagnósis de los residuos

```
# diagnósis de los residuos ----

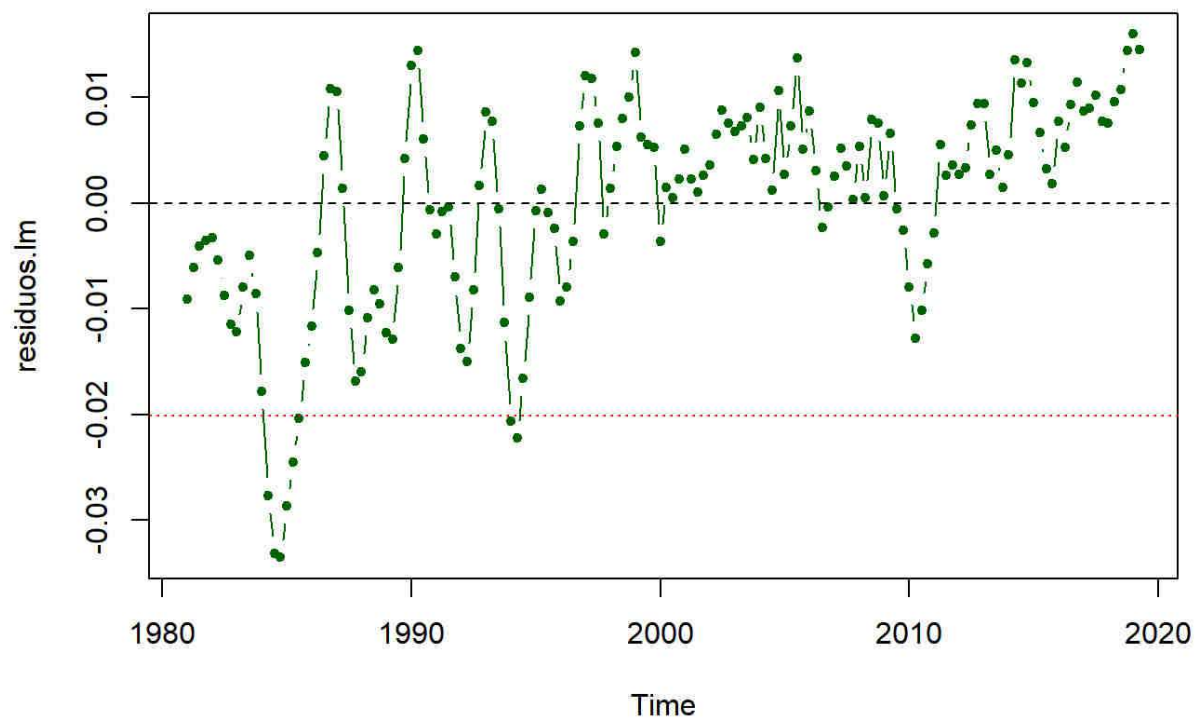
residuos.lm<-residuals(modelo.lm)

plot(residuos.lm, type="b", col="darkgreen",pch=20)
abline(h=0,lty=2, col="grey2" )
abline(h=2*sd(residuos.lm),lty=3, col="red" )
abline(h=-2*sd(residuos.lm),lty=3, col="red" )

# para seleccionar atípicos en pantalla
#t<-locator(type="p", col="green",pch= 19)
#t

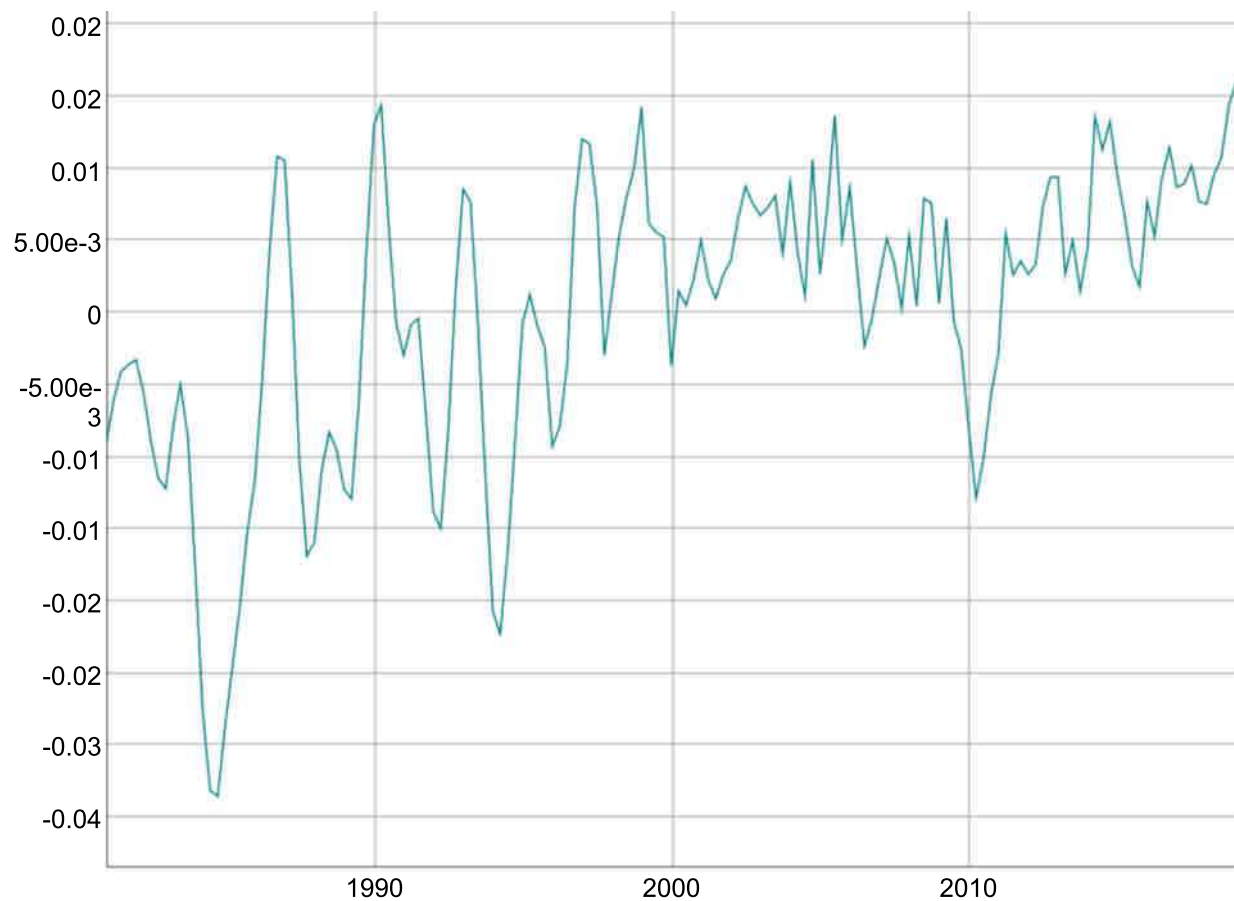
# También se pueden identificar así
#identify(x = residuos.lm, labels = time(residuos.lm))

# o con la library(dygraphs)
library(dygraphs)
```



```
dygraph(residuos.lm)
```

```
## Registered S3 method overwritten by 'xts':
##   method      from
##   as.zoo.xts zoo
```



## 7 Tarea Punto 2: ¿A qué ritmo debe crecer la economía para empezar a generar empleo?

```
# Tarea Punto 2: ¿ a qué ritmo debe crecer la economía para empezar a generar empleo? ----
```

```
coef(modelo.lm)
```

```
## (Intercept)      T4pib.ts  
## -0.01744322  1.24371091
```

```
coef1<-coef(modelo.lm)[1]  
coef2<-coef(modelo.lm)[2]
```

```
tasaobj<--coef1/coef2
```

```
cat(paste0("La tasa de crecimiento necesaria para empezar a generar empleo es ", round(tasaobj,digit  
s = 4), "\n"))
```

```
## La tasa de crecimiento necesaria para empezar a generar empleo es 0.014
```

```
# Continuamos con la diagnosis del modelo
```

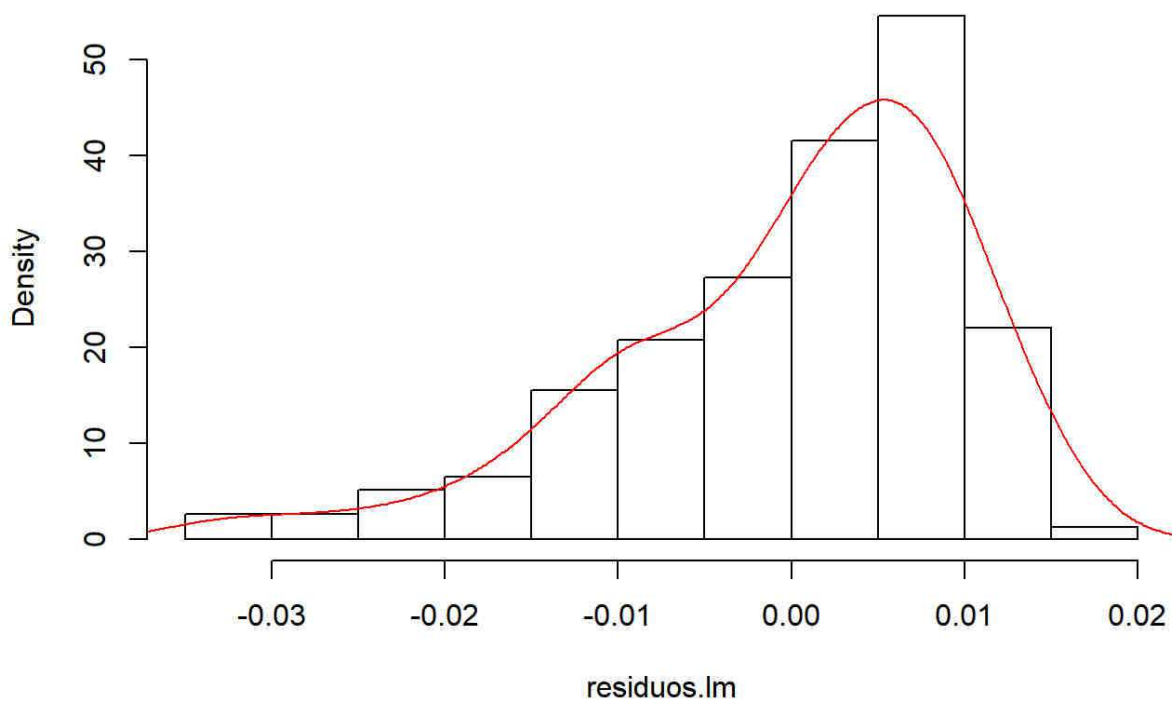
## 8 Continuamos con la diagnosis del modelo

### 8.1 Test de Normalidad sobre los residuos

```
# Test de Normalidad ----
```

```
hist(residuos.lm, freq=FALSE)  
lines(density(residuos.lm), col=2)
```

## Histogram of residuos.lm



```
library(tseries)
```

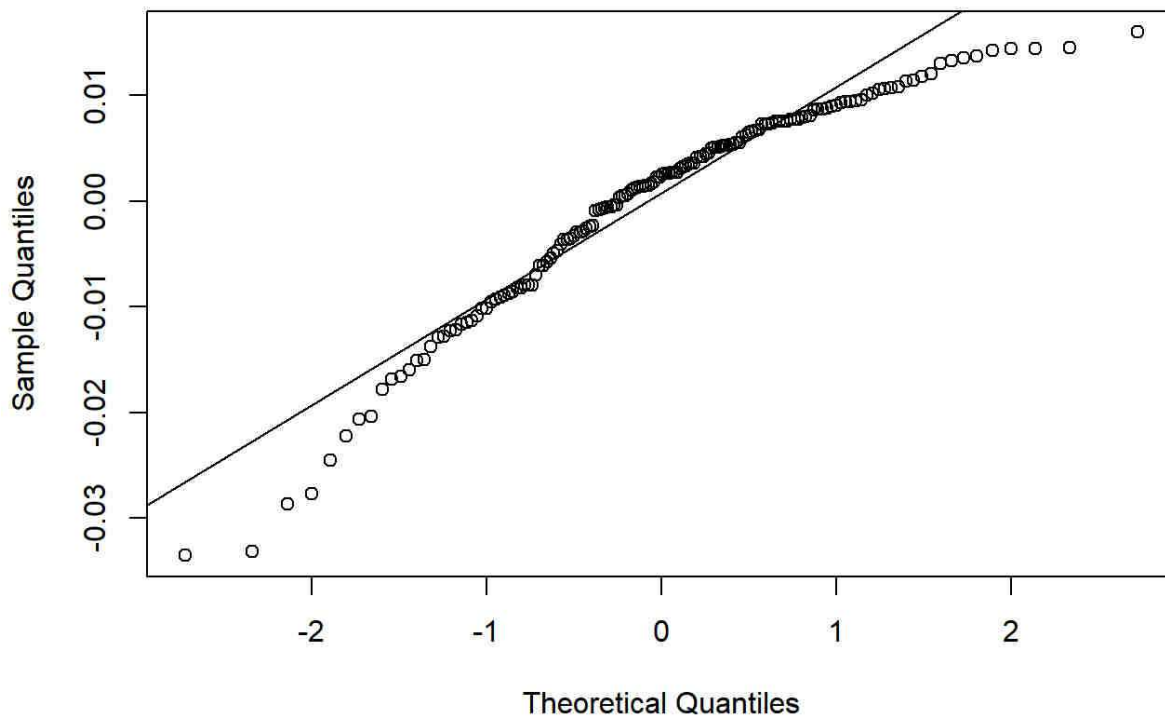
```
## Registered S3 method overwritten by 'quantmod':  
##   method      from  
##   as.zoo.data.frame zoo
```

```
jarque.bera.test(residuos.lm) ##H0:Normalidad
```

```
##  
## Jarque Bera Test  
##  
## data:  residuos.lm  
## X-squared = 31.165, df = 2, p-value = 1.708e-07
```

```
qqnorm(residuos.lm)  
qqline(residuos.lm)
```

## Normal Q-Q Plot



```
#Test de Normalidad de Shapiro-Wilk, H0:Normalidad
shapiro.test(residuos.lm)
```

```
##
##  Shapiro-Wilk normality test
##
## data:  residuos.lm
## W = 0.93063, p-value = 8.246e-07
```

```
# Ojo quizás sería más conveniente utilizar los residuos estudentizados para comprobar la hipótesis
# de normalidad
# los residuos estudentizados, son los residuos que se tipifican (dividen) no
# por la varianza del conjunto de los residuos, sino que cada residuo i
# se divide por la varianza residual calculada con toda la muestra
# excepto la propia observación i

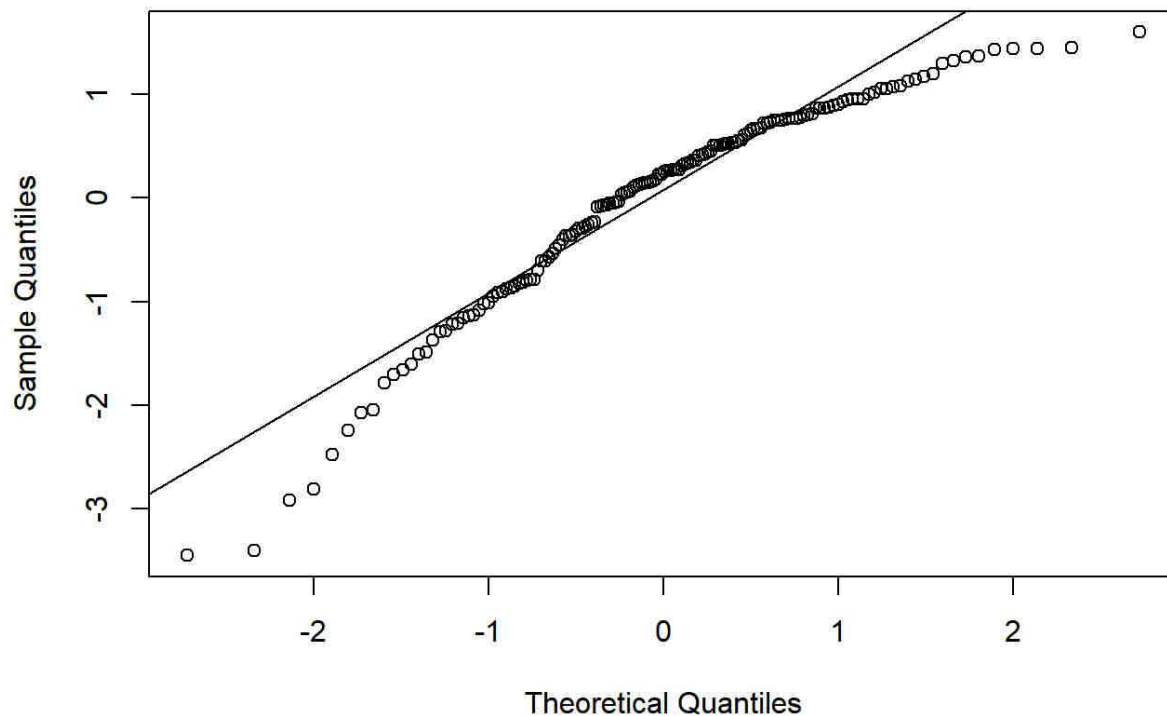
library(MASS)
residuos.lm.estund<-studres(modelo.lm)

jarque.bera.test(residuos.lm.estund) ##H0:Normalidad
```

```
##
##  Jarque Bera Test
##
## data:  residuos.lm.estund
## X-squared = 35.23, df = 2, p-value = 2.238e-08
```

```
qqnorm(residuos.lm.estund)
qqline(residuos.lm.estund)
```

Normal Q-Q Plot



```
shapiro.test(residuos.lm.estund)
```

```
##  
## Shapiro-Wilk normality test  
##  
## data:  residuos.lm.estund  
## W = 0.92768, p-value = 5.129e-07
```

## 8.2 Contraste de Hipótesis

```
# Contraste de Hipótesis ----
```

## 8.3 Contraste de Especificación de Ramsey

```
# Contraste de Especificación de Ramsey (Ho: el modelo está bien especificado)  
# El test RESET añade diferentes formas polinómicas en el ajuste OLS  
# para detectar diferentes tipos de mala-especificación general
```

```
library(lmtest)
```

```
## Loading required package: zoo
```

```
##  
## Attaching package: 'zoo'
```



```
## The following objects are masked from 'package:base':
##
##   as.Date, as.Date.numeric
```

```
resettest(modelo.lm)
```

```
##
## RESET test
##
## data:  modelo.lm
## RESET = 8.2375, df1 = 2, df2 = 150, p-value = 0.0004033
```

## 8.4 Contraste de hipótesis lineales sobre los coeficientes

```
#contrate de hipótesis lineales sobre los coeficientes
```

```
# ejemplo: librería car
#H0 : b0 = 0; b3 + B4 = 1

#unrestricted <- lm(y~x1+x2+x3+x4)
#rhs <- c(0,1)
#hm <- rbind(c(1,0,0,0,0),c(0,0,0,1,1))
#car::linear.hypothesis(unrestricted,hm,rhs)

library(car)
```

```
## Loading required package: carData
```

```
#Nosotros vamos a contrastar si la pendiente es igual a 1
#H0: b1=1
rhs<-1
hm<-c(0,1)
#car::linear.hypothesis(modelo.lm,hm,rhs)
car::linearHypothesis(modelo.lm,hm,rhs)
```

```
## Linear hypothesis test
##
## Hypothesis:
## T4pib.ts = 1
##
## Model 1: restricted model
## Model 2: T4ocu.ts ~ T4pib.ts
##
##   Res.Df      RSS Df Sum of Sq      F      Pr(>F)
## 1     153 0.020044
## 2     152 0.015503   1 0.0045418 44.531 4.395e-10 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
# ¿es la tasa de crecimiento necesaria para generar empleo estadísticamente =2% ?
# 0.02=-b0/b1
# H0: 0.02*b1+b0=0
rhs<-0
hm<-c(1,0.02)
# car::linear.hypothesis(modelo.lm,hm,rhs)
car::linearHypothesis(modelo.lm,hm,rhs)
```

```
## Linear hypothesis test
##
## Hypothesis:
## (Intercept) + 0.02 T4pib.ts = 0
##
## Model 1: restricted model
## Model 2: T4ocu.ts ~ T4pib.ts
##
##      Res.Df      RSS Df Sum of Sq      F      Pr(>F)
## 1      153 0.023832
## 2      152 0.015503   1 0.0083292 81.665 6.925e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

## 8.5 Contraste sobre un único parámetro

```
# Contraste sobre un único parámetro
# Coeficientes
coef(modelo.lm)
```

```
## (Intercept)      T4pib.ts
## -0.01744322  1.24371091
```

```
# matriz de Varianzas y covarianzas del vector de parámetros estimados
vcov(modelo.lm)
```

```
##              (Intercept)      T4pib.ts
## (Intercept)  1.381849e-06 -3.097986e-05
## T4pib.ts     -3.097986e-05  1.333793e-03
```

```
# varianzas de cada parámetro individual
diag(vcov(modelo.lm))
```

```
## (Intercept)      T4pib.ts
## 1.381849e-06 1.333793e-03
```

```
# desviación típica de cada parámetro individual
sqrt(diag(vcov(modelo.lm)))
```

```
## (Intercept)      T4pib.ts
## 0.001175521 0.036521136
```

```
summary(modelo.lm)
```

```
##
## Call:
## lm(formula = modelo, data = datokun, na.action = NULL)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.033511 -0.005997  0.002390  0.007533  0.015968
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -0.017443   0.001176  -14.84  <2e-16 ***
## T4pib.ts     1.243711   0.036521   34.05  <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.0101 on 152 degrees of freedom
## Multiple R-squared:  0.8841, Adjusted R-squared:  0.8834
## F-statistic: 1160 on 1 and 152 DF,  p-value: < 2.2e-16
```

```
# estadístico t H0=0
t_test<-coef(modelo.lm)[1]/sqrt(vcov(modelo.lm)[1,1])

p_valor<-2*(pt(-abs(t_test),df=grLibtd,lower.tail = TRUE)) #P(|t_tes|>0)
p_valor
```

```
## (Intercept)
## 2.315177e-31
```

```
t_test<-coef(modelo.lm)[2]/sqrt(vcov(modelo.lm)[2,2])
t_test
```

```
## T4pib.ts
## 34.05455
```

```
p_valor<-2*(pt(-abs(t_test),df=grLibtd,lower.tail = TRUE)) #P(|t_tes|>0)
p_valor
```

```
## T4pib.ts
## 5.024061e-73
```

```
salida.lm$coefficients
```

```
##              Estimate Std. Error  t value    Pr(>|t|)
## (Intercept) -0.01744322 0.001175521 -14.83871 2.315177e-31
## T4pib.ts     1.24371091 0.036521136  34.05455 5.024061e-73
```

```
# ¿es la pendiente estadísticamente igual a uno?
```

## 8.6 Contraste de hipótesis por Sumas Residuales

```
# Contraste por Sumas Residuales
```

```
modeloRes<-modelo.lm  
summary(modeloRes)
```

```
##  
## Call:  
## lm(formula = modelo, data = datokun, na.action = NULL)  
##  
## Residuals:  
##      Min       1Q   Median       3Q      Max   
## -0.033511 -0.005997  0.002390  0.007533  0.015968   
##  
## Coefficients:  
##              Estimate Std. Error t value Pr(>|t|)      
## (Intercept) -0.017443   0.001176  -14.84  <2e-16 ***   
## T4pib.ts     1.243711   0.036521   34.05  <2e-16 ***   
## ---  
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1  
##  
## Residual standard error: 0.0101 on 152 degrees of freedom  
## Multiple R-squared:  0.8841, Adjusted R-squared:  0.8834   
## F-statistic: 1160 on 1 and 152 DF,  p-value: < 2.2e-16
```

```
modeloSinR<-lm(T4ocu.ts ~ T4pib.ts + I(T4pib.ts^2), data=datokun)  
summary(modeloSinR)
```

```
##  
## Call:  
## lm(formula = T4ocu.ts ~ T4pib.ts + I(T4pib.ts^2), data = datokun)  
##  
## Residuals:  
##      Min       1Q   Median       3Q      Max   
## -0.032642 -0.005553  0.001926  0.007530  0.016692   
##  
## Coefficients:  
##              Estimate Std. Error t value Pr(>|t|)      
## (Intercept) -0.018129   0.001317  -13.76  <2e-16 ***   
## T4pib.ts     1.206729   0.048639   24.81  <2e-16 ***   
## I(T4pib.ts^2) 1.491126   1.297047    1.15    0.252      
## ---  
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1  
##  
## Residual standard error: 0.01009 on 151 degrees of freedom  
## Multiple R-squared:  0.8851, Adjusted R-squared:  0.8836   
## F-statistic: 581.7 on 2 and 151 DF,  p-value: < 2.2e-16
```

```
anova(modeloRes,modeloSinR)
```

```
## Analysis of Variance Table
##
## Model 1: T4ocu.ts ~ T4pib.ts
## Model 2: T4ocu.ts ~ T4pib.ts + I(T4pib.ts^2)
##   Res.Df      RSS Df Sum of Sq    F Pr(>F)
## 1     152 0.015503
## 2     151 0.015368  1 0.00013451 1.3217 0.2521
```

## 9 Predicción en el modelo lineal

```
# Predicciones -----
# Tarea punto 3_ PRedicciones
# debeis utilizar el comando: predict()
# pero primero necesitaréis predicciones de t4pib.ts

#ejemplo: supongamos que el PIB crece un 5% el próximo año ¿cuanto crecerían los ocupados?

newTocu=sum(coef(modelo.lm)*c(1,0.05))
print(newTocu)
```

```
## [1] 0.04474233
```

```
# ¿intervalos de confianza?
# supongamos que este año se crecerá al 2.2% y el año que viene 2.8%
nuevosdatos= data.frame(T4pib.ts=c(0.022,0.028)) # supongamos que
prediccion<-predict(modelo.lm,newdata = nuevosdatos, se.fit = TRUE, interval="prediction", level=0.9
5)
prediccion
```

```
## $fit
##           fit           lwr           upr
## 1 0.009918422 -0.010099166 0.02993601
## 2 0.017380687 -0.002639667 0.03740104
##
## $se.fit
##           1           2
## 0.0008150406 0.0008322685
##
## $df
## [1] 152
##
## $residual.scale
## [1] 0.01009909
```

```
# último dato de Ocupados
ocu2018<-okun$ocu.ts[time(okun$ocu.ts)==2018.75]

ocu2019<-ocu2018*(1+prediccion$fit[1,1])
ocu2020<-ocu2018*(1+prediccion$fit[1,1])*(1+prediccion$fit[2,1])

cat(paste0("En 2019 se genererán ", round(ocu2019-ocu2018, digits = 1), " miles de personas", "\n"))
```

```
## En 2019 se genererán 184.4 miles de personas
```

```
cat(paste0("Entre 2019 y 2020 se genererán ", round(ocu2020-ocu2018, digits = 1), " miles de personas", "\n"))
```

```
## Entre 2019 y 2020 se genererán 510.6 miles de personas
```

```
# ¿a qué ritmo debería crecerse para que se generen 2 millones de puestos de trabajo?  
# habría que despejar de aquí  
#  $ocu2018 * (1 + tasaobj)^2 - ocu2018 = 2000$   
 $tasaobj = (2000 / (ocu2018 + 1))^{0.5} - 1$   
print(tasaobj)
```

```
## [1] 0.05242434
```

## 10 Test de Cambio Estructural

```
#Test de Cambio Estructural----
```

Los test de cambio estructural sirven para contrastar la hipótesis nula de que no ha habido cambios en los parámetros del modelo a lo largo de toda la muestra, que los parámetros del modelo han sido siempre los mismos

```
## H0: AUSENCIA DE CAMBIO ESTRUCTURAL  
library(strucchange)
```

```
## Loading required package: sandwich
```

### 10.1 Test de Chow de Cambio Estructural

```
# Test de Cambio estructural Chow  
# Para contrastar la H0 de que hay dos modelos diferentes uno antes y otro después de 1997  
Chowtest<-Fstats(modelo.lm, from=c(1997,1), to=c(1997,1), data=datokun)  
Chowtest$Fstats
```

```
##           Qtr1  
## 1997 84.26371
```

```
#sctest(Chowtest, type="expF")  
sctest(Chowtest, type="aveF")
```

```
##  
## Chow test  
##  
## data: Chowtest  
## ave.F = 84.264, p-value < 2.2e-16
```

```
sctest(Chowtest, type="supF")
```

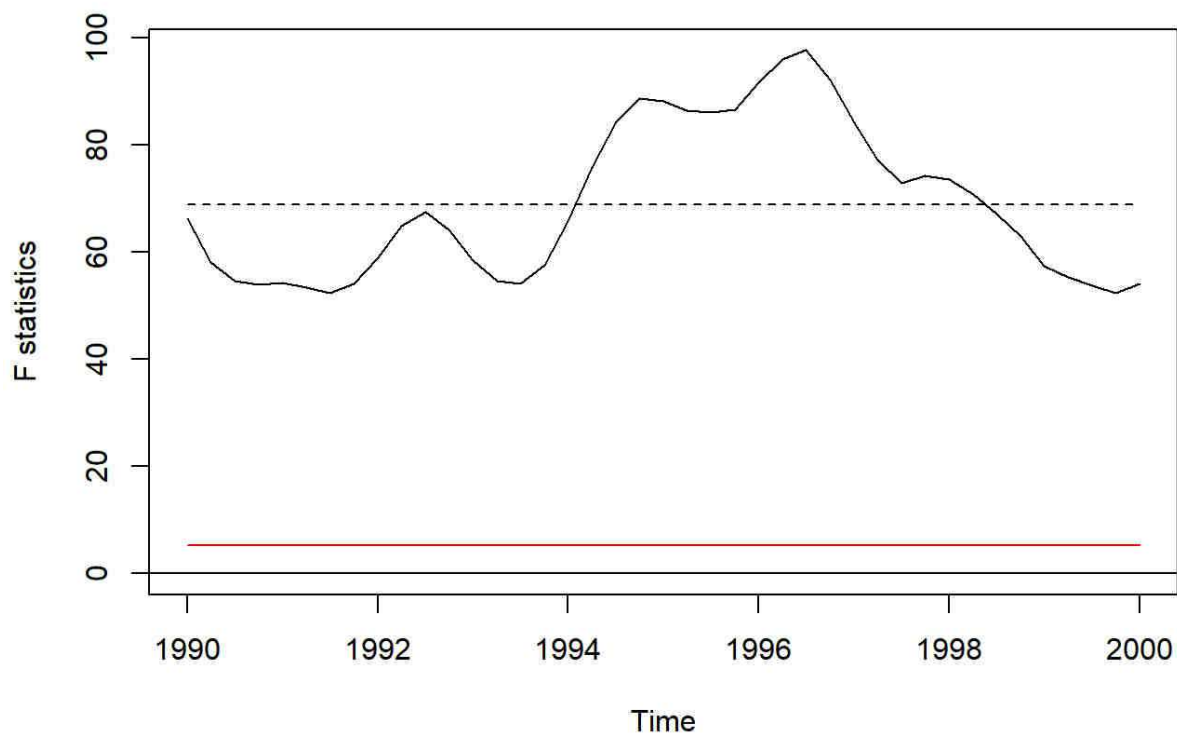
```
##
## Chow test
##
## data: Chowtest
## sup.F = 84.264, p-value < 2.2e-16
```

PARa realizar el test con múltiples puntos de corte (cuando se desconoce el punto en el que puede haberse producido el cambio estructural)

```
Chowtest<-Fstats(modelo, from=c(1990,1), to=c(2000,1), data=datokun)
Chowtest$Fstats
```

```
##      Qtr1    Qtr2    Qtr3    Qtr4
## 1990 66.27900 58.17136 54.62999 54.01970
## 1991 54.26544 53.47155 52.39308 54.18043
## 1992 59.08142 64.95482 67.49634 64.24026
## 1993 58.41123 54.64663 54.19603 57.67809
## 1994 65.73192 75.94880 84.36931 88.77427
## 1995 88.18821 86.44114 86.10575 86.73204
## 1996 91.65844 96.00438 97.74731 92.25166
## 1997 84.26371 77.23248 72.97241 74.36825
## 1998 73.58810 70.90796 67.23974 62.97472
## 1999 57.38263 55.41084 53.79414 52.42556
## 2000 54.10319
```

```
plot(Chowtest, aveF=TRUE)
locator(1,type = "p", col=2, pch=19)
```



```
sctest(Chowtest, type="aveF")
```

```
##  
## aveF test  
##  
## data: Chowtest  
## ave.F = 68.896, p-value < 2.2e-16
```

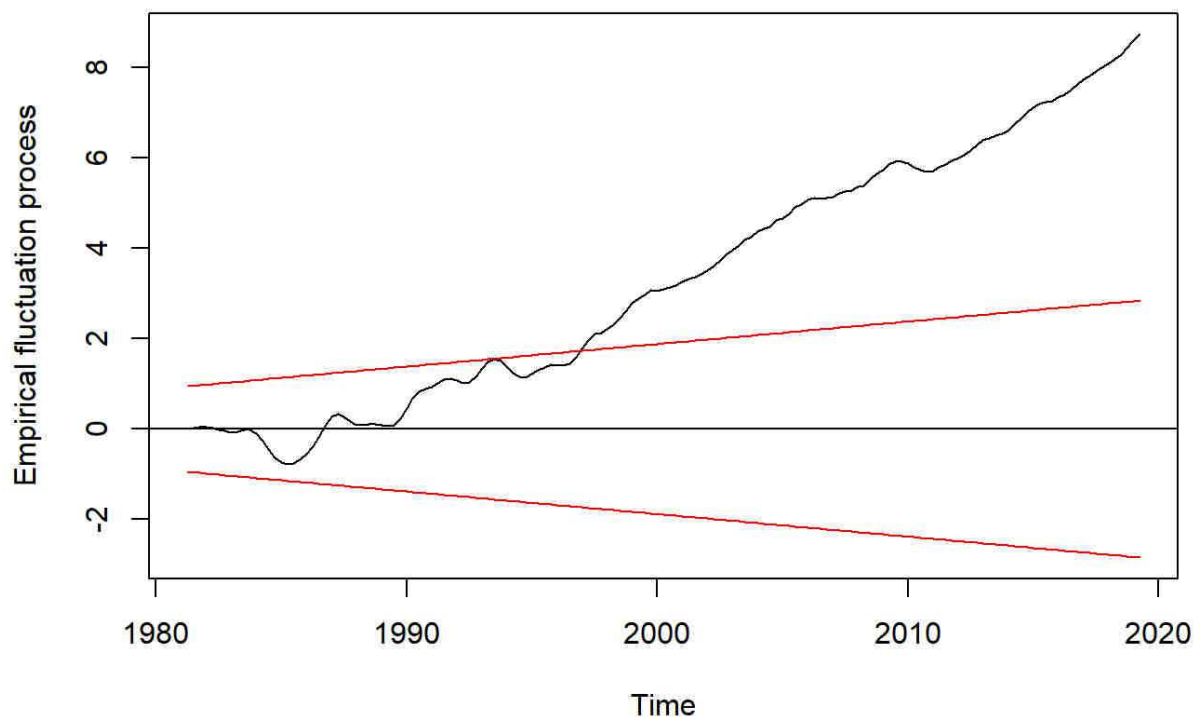
```
sctest(Chowtest, type="supF")
```

```
##  
## supF test  
##  
## data: Chowtest  
## sup.F = 97.747, p-value < 2.2e-16
```

## 10.2 Test de Cambio Estructural mediante mínimos Cuadrados Recursivos

```
# Mínimos cuadrados recursivos  
  
# Diferencias entre los residuos  
testestbr<-efp(modelo, type="Rec-CUSUM", data = datokun, h=0.2) # Este utiliza el error de predicción,  
# MISMO QUE EN EViews  
plot(testestbr)
```

**Recursive CUSUM test**





```
# para identificar puntos de cambio
#locator(10, type = "p" ,col=2, pch=19)

#plot(testestbr$process)
# contraste para saber si se mantiene entre bandas
sctest(testestbr)
```

```
##
## Recursive CUSUM test
##
## data: testestbr
## S = 2.9142, p-value = 3.553e-15
```

Estos Test de cambio estructural presentan muchas posibilidades, por lo que dejamos para la posterior práctica el uso específico de estos test de Cambio Estructural

```
# RESUMEN CAMBIO ESTRUCTURAL
#####33
datokun1<-ts.intersect(T4pib.ts,T4ocu.ts,dframe = FALSE)
datokun2<-window(datokun1,start=2009, end=tiempo_max, frequency=4)
Chowtest<-Fstats(modelo, from=c(2010,1), to=c(2012,4), data=datokun2)
Chowtest$Fstats
```

```
##           Qtr1      Qtr2      Qtr3      Qtr4
## 2010  5.308436 20.011604 40.887025 57.403695
## 2011 69.219183 50.715127 48.285523 46.275468
## 2012 47.564548 49.071966 46.118334 43.491371
```

```
sctest(Chowtest, type="expF")
```

```
##
## expF test
##
## data: Chowtest
## exp.F = 32.128, p-value = 8.915e-12
```

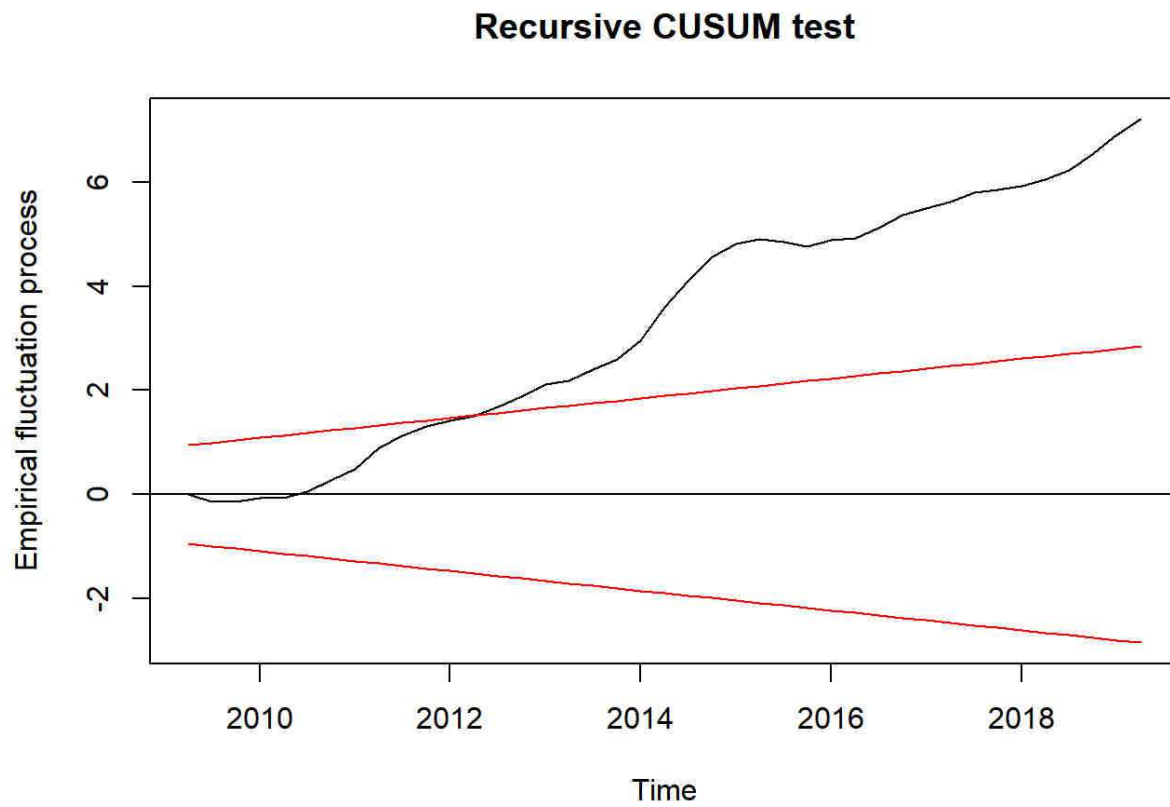
```
sctest(Chowtest, type="aveF") # tomo este como bueno
```

```
##
## aveF test
##
## data: Chowtest
## ave.F = 43.696, p-value = 2.314e-11
```

```
sctest(Chowtest, type="supF")
```

```
##
## supF test
##
## data: Chowtest
## sup.F = 69.219, p-value = 6.631e-14
```

```
testestbr<-efp(modelo, type="Rec-CUSUM", data = datokun2, h=0.2) # ESTE ES EL MISMO QUE EN EVIEWS
plot(testestbr)
```



```
sctest(testestbr)
```

```
##
## Recursive CUSUM test
##
## data: testestbr
## S = 2.4014, p-value = 1.911e-10
```

```
#####33
```



Proyecto de innovación  
“Innova Docencia” de la UCM  
Convocatoria 2018-2019  
Proyecto número 275

**R-adaptación de la asignatura de métodos econométricos en economía y finanzas del  
Grado de Estadística Aplicada**

Facultad de Estudios Estadísticos  
Departamento de Economía Aplicada, Pública y Política

**Título**

***Práctica 3 : La Ley de Okun: Test de Cambio Estructural, ¿han sido  
“eficaces” las reformas del mercado de trabajo?***

**Correspondiente al**

**Capítulo 2 REGRESIÓN LINEAL**

- El Modelo Lineal General (MLG)
- Hipótesis del modelo
- Estimación por Mínimos Cuadrados Ordinarios (MCO)
- Inferencia y predicción en el MLG
- Contrastes de cambio estructural
- Multicolinealidad y errores de medida

***Práctica 2 : La Ley de Okun: ¿a qué ritmo debe crecer la Economía para  
comenzar a generar empleo***

***Práctica 3 : La Ley de Okun: Test de Cambio Estructural, ¿han sido  
“eficaces” las reformas del mercado de trabajo?***

***Práctica 4 : Ley de Okun con datos de la UE desde la web de Eurostat***

***Práctica 5 : La ecuación de salarios y la discriminación por razón de género***

# P3. - Ley de Okun: test de Cambio Estructural

## Asignatura Métodos Econométricos en Economía y Finanzas

Profesores  
Loren Escot, Alicia Pérez-Alonso y Julio E. Sandubete

Dpto Economía Aplicada, Pública y Política  
Facultad Estudios Estadísticos - Universidad Complutense de Madrid

Curso académico 2018/2019

### Abstract

En esta práctica se muestra cómo aplicar la librería `strucchange` para realizar test de Cambio estructural, esto es, contrastes de la hipótesis nula de que no ha habido cambios en los parámetros del modelo a lo largo de toda la muestra, que los parámetros del modelo han sido siempre los mismos.

Se utilizarán estos test de cambio estructural para mostrar cómo pueden utilizarse para contrastar si las distintas reformas laborales han sido eficaces, esto es, permiten una mayor generación de empleo con la misma tasa de crecimiento y/o un menor ritmo de crecimiento del PIB para empezar a generar empleo

- 1 Librerías necesarias
- 2 Ley de Okun
- 3 Test de Cambio Estructural,  $H_0$ : AUSENCIA DE CAMBIO ESTRUCTURAL
  - 3.1 Test de Cambio estructural sobre la serie simulada
    - 3.1.1 Test de Cambio estructural Chow
    - 3.1.2 Test de Cambio estructural Mínimos Cuadrados recursivos
- 4 Tarea punto 4\_ Cambio estructural: ¿Han sido eficaces las reformas que ha sufrido el mercado de trabajo?
- 5 Ahora Vosotros
  - 5.1 Contrastad si ha habido cambio en 2006
  - 5.2 Contrastad si ha habido cambio en 2009
  - 5.3 Contrastad si ha habido cambio 2010
  - 5.4 Contrastad si ha habido cambio 2012

## 1 Librerías necesarias

```
library(stargazer) # para hacer tablas
```

```
##  
## Please cite as:
```

```
## Hlavac, Marek (2018). stargazer: Well-Formatted Regression and Summary Statistics Tables.
```

```
## R package version 5.2.2. https://CRAN.R-project.org/package=stargazer
```

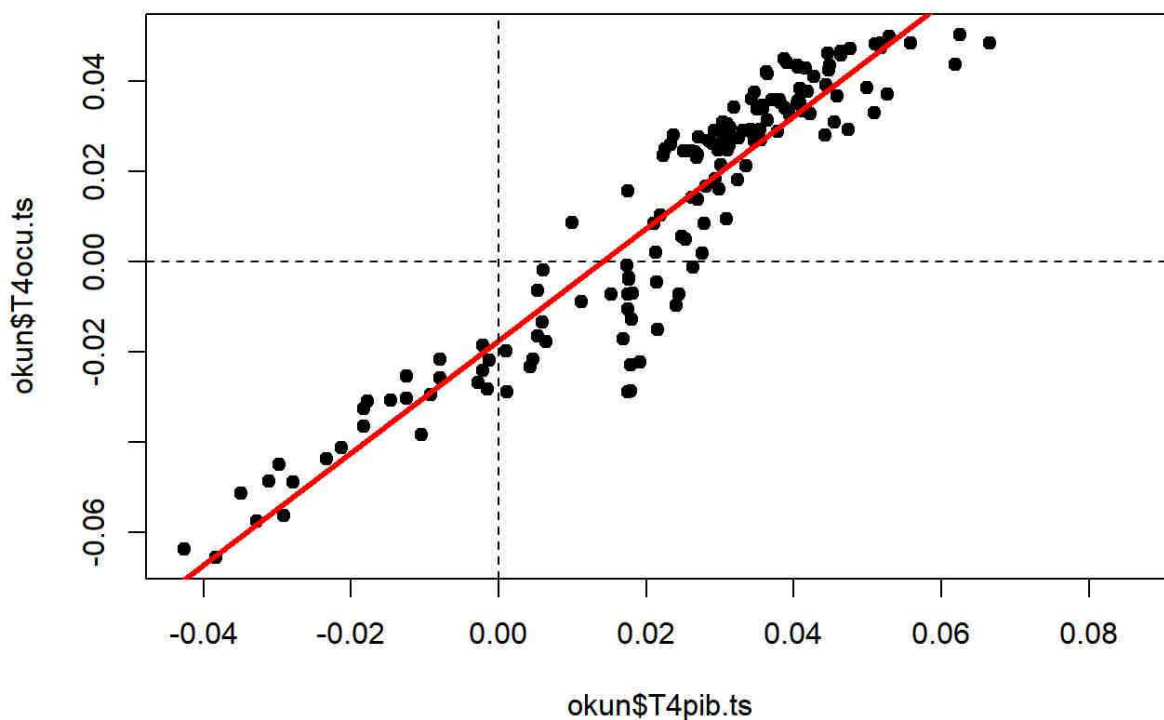
Carga de datos

```
load(file="okundat.Rdata")
# ultimo dato disponible
tiempo_max<-max(okun$tiempo)
print(tiempo_max)
```

```
## [1] 2019.25
```

## 2 Ley de okun

```
plot(okun$T4pib.ts, okun$T4ocu.ts, pch=19)
abline(lm(T4ocu.ts ~ T4pib.ts, data=okun), col=2,lwd=3)
abline(h=0,lty=2, col="grey2")
abline(v=0,lty=2, col="grey2")
```



```
modelo<-T4ocu.ts ~ T4pib.ts
```

```
# para poder trabajar con el modelo de regresión cuando trabajamos con series temporales
# conviene asegurarse de que todas las series tienen el mismo rango temporal
```

```
T4pib.ts<-window(okun[,5], start=1981, end=tiempo_max)
T4ocu.ts<-window(okun[,7], start=1981, end=tiempo_max)
```

```
datokun<-ts.intersect(T4pib.ts,T4ocu.ts,dframe = TRUE)
```

```
modelo.lm<-lm(modelo, data=datokun, na.action=NULL) #ojo, na.action=NULL para las series
```

```
stargazer(modelo.lm, type="text")
```

```
##
## =====
##                      Dependent variable:
##                      -----
##                      T4ocu.ts
## -----
## T4pib.ts                1.244***
##                        (0.037)
##
## Constant                -0.017***
##                        (0.001)
## -----
## Observations              154
## R2                      0.884
## Adjusted R2              0.883
## Residual Std. Error      0.010 (df = 152)
## F Statistic              1,159.712*** (df = 1; 152)
## =====
## Note:                    *p<0.1; **p<0.05; ***p<0.01
```

### 3 Test de Cambio Estructural, H0: AUSENCIA DE CAMBIO ESTRUCTURAL

```
#Test de Cambio Estructural----
```

```
## H0: AUSENCIA DE CAMBIO ESTRUCTURAL
library(strucchange)
```

```
## Loading required package: zoo
```

```
##
## Attaching package: 'zoo'
```

```
## The following objects are masked from 'package:base':
##
##      as.Date, as.Date.numeric
```

```
## Loading required package: sandwich
```

Ejemplo de serie simulada con cambio estructural

```
# Pongamos un ejemplo

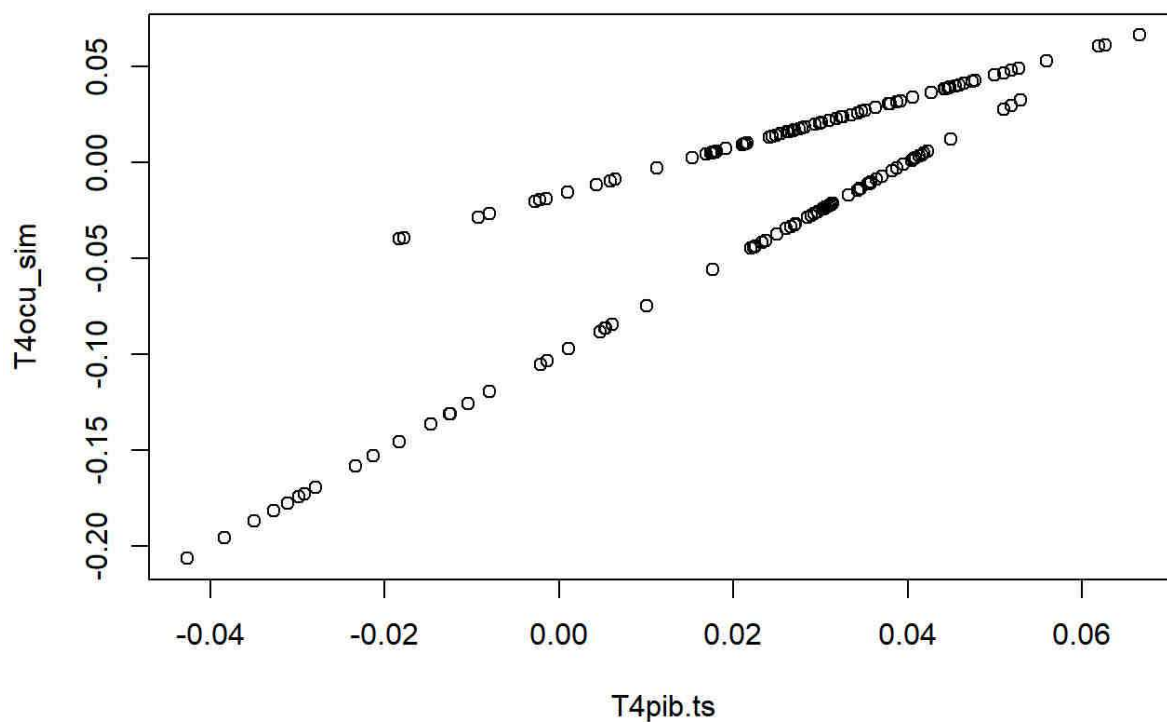
#supongamos que tenemos en realidad dos muestras diferentes generadas con dos modelos diferentes

T4ocu_sim1<-(-0.017+ 1.25*window(T4pib.ts, start=1981, end=2000))
T4ocu_sim2<-(-0.10 + 2.50*window(T4pib.ts, start=2000.25, end=tiempo_max))

T4ocu_sim<-ts(c(T4ocu_sim1,T4ocu_sim2), start=1981, frequency = 4)

# No existe una recta de regresión sino dos una desde 1981 hasta 2000 y
# otra a partir de 2000 (por construcción)

plot(T4ocu_sim~T4pib.ts)
```



```
# Vamos a ver como detectarlo
datokunsim<-ts.union(T4ocu.ts=T4ocu_sim,T4pib.ts, dframe=TRUE)

modelo.sim.lm<-lm(modelo, data=datokunsim, na.action=NULL)

stargazer(modelo.sim.lm, type="text")
```

```
##
## =====
##                      Dependent variable:
##                      -----
##                      T4ocu.ts
## -----
## T4pib.ts                2.318***
##                        (0.104)
##
## Constant                -0.072***
##                        (0.003)
## -----
## Observations              154
## R2                       0.766
## Adjusted R2              0.764
## Residual Std. Error      0.029 (df = 152)
## F Statistic              496.873*** (df = 1; 152)
## =====
## Note:                    *p<0.1; **p<0.05; ***p<0.01
```

## 3.1 Test de Cambio estructural sobre la serie simulada

### 3.1.1 Test de Cambio estructural Chow

```
# Test de Cambio estructural Chow
```

```
Chowtest<-Fstats(modelo.sim.lm, from=c(2000,1), to=c(2000,1), data=datokunsim)
Chowtest$Fstats
```

```
##           Qtr1
## 2000 6.191469e+31
```

```
sctest(Chowtest, type="expF")
```

```
##
## expF test
##
## data: Chowtest
## exp.F = Inf, p-value < 2.2e-16
```

```
sctest(Chowtest, type="aveF")
```

```
##
## Chow test
##
## data: Chowtest
## ave.F = 6.1915e+31, p-value < 2.2e-16
```

```
sctest(Chowtest, type="supF")
```

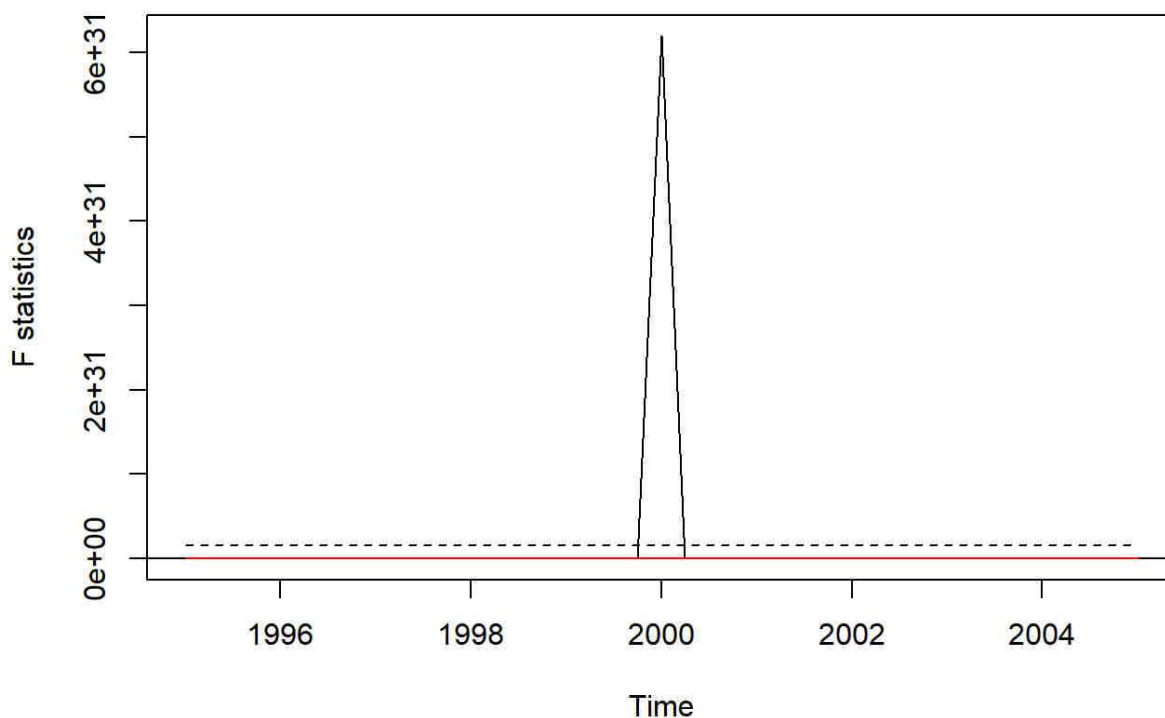


```
##
## Chow test
##
## data: Chowtest
## sup.F = 6.1915e+31, p-value < 2.2e-16
```

```
# si no supiera a priori en qué año se produce el cambio estructural puedo probar en
# varios puntos de corte
Chowtest<-Fstats(modelo, from=c(1995,1), to=c(2005,1), data=datokunsim)
Chowtest$Fstats
```

```
##          Qtr1          Qtr2          Qtr3          Qtr4
## 1995 7.080640e+02 7.643765e+02 8.415321e+02 9.433477e+02
## 1996 1.087803e+03 1.288386e+03 1.482660e+03 1.817015e+03
## 1997 2.162498e+03 2.555676e+03 3.109210e+03 3.357569e+03
## 1998 3.730754e+03 4.326505e+03 4.994176e+03 6.567102e+03
## 1999 1.056853e+04 1.513715e+04 2.772624e+04 1.131202e+05
## 2000 6.191469e+31 6.903153e+04 3.272462e+04 2.106699e+04
## 2001 1.222559e+04 7.530208e+03 5.345925e+03 3.828655e+03
## 2002 2.764442e+03 2.147166e+03 1.667400e+03 1.365841e+03
## 2003 1.198992e+03 1.061818e+03 9.578176e+02 8.873786e+02
## 2004 8.078665e+02 7.452859e+02 7.031818e+02 6.518709e+02
## 2005 6.192354e+02
```

```
plot(Chowtest, aveF=TRUE)
```



```
# En este caso es evidente que el cambio se produce en el año 2000
```

```
#locator(1,type = "p", col=2, pch=19)
```

```
sctest(Chowtest, type="aveF")
```

```
##  
## aveF test  
##  
## data: Chowtest  
## ave.F = 1.5101e+30, p-value < 2.2e-16
```

```
sctest(Chowtest, type="supF")
```

```
##  
## supF test  
##  
## data: Chowtest  
## sup.F = 6.1915e+31, p-value < 2.2e-16
```

## 3.1.2 Test de Cambio estructural Mínimos Cuadrados recursivos

```
# Mínimos cuadrados recursivos
```

```
# Existen diferentes test según como se midan las diferencias
```

- # Rec: Errores de predicción (estandarizados)
- # OLS: Residuos recursivos (estandarizados)
- # Diferencia entre coeficientes

```
# Y como se midan esas diferencias
```

- # CUSUM: Suma Acumulada
- # MOSUM: Media Movil Acumulada

```
# con esto tenemos 6 tipos diferentes
```

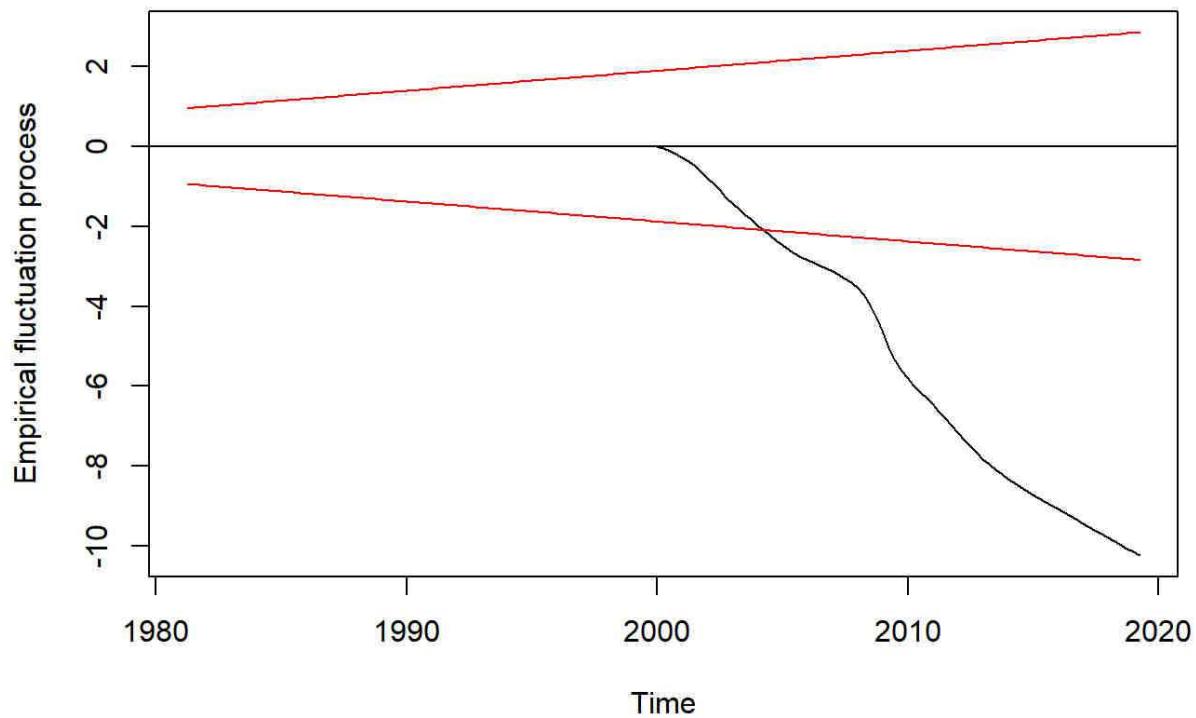
- # 1. type="Rec-CUSUM": CUSUM Errores de predicción ### Utilizar ESTE para la práctica
- # 2. type="OLS-CUSUM": CUSUM Residuos
- # 3. type="Rec-MOSUM": Media Movil Errores de predicción
- # 4. type="OLS-MOSUM": Media Movil Residuos
- # 5. type="RE": DiferenciaAcumulada coeficientes
- # 6. type="ME": Media movil diferencia entre coeficientes

```
# Se calculan con la función "efp"
```

```
# Diferencias entre los residuos
```

```
testestbr<-efp(modelo, type="Rec-CUSUM", data = datokunsim, h=0.2) # Este utiliza el error de predic  
ci?n, MISMO QUE EN EVIEWS  
plot(testestbr)
```

## Recursive CUSUM test



```
#locator(1, type = "p" ,col=2, pch=19)
#plot(testestbr$process)

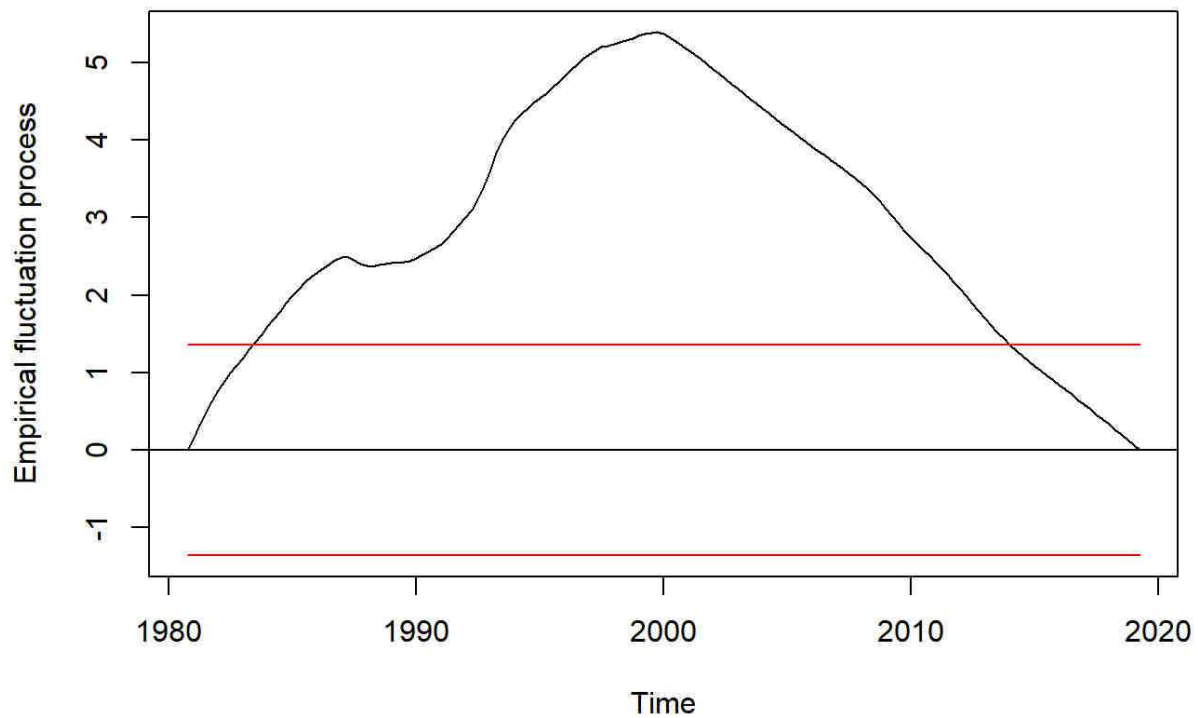
# Nótese que se sale de bandas y por tanto rechazáramos la H0 de ausencia de cambio estructural
# Nótese igualmente que el cambio estructural sabemos que se produce a partir del año 2000
# (por contrucción de las series simuladas) y que efectivamente a partir del año 2000 se
# produce un CAMBIO DE TENDENCIA, aunque se salga de bandas más adelante

# contraste para saber si se mantiene entre bandas
sctest(testestbr)
```

```
##
## Recursive CUSUM test
##
## data: testestbr
## S = 3.4094, p-value < 2.2e-16
```

```
# otras opciones
testestbr<-efp(modelo, type="OLS-CUSUM", data = datokunsim) # este utiliza ls residuos
plot(testestbr)
```

## OLS-based CUSUM test

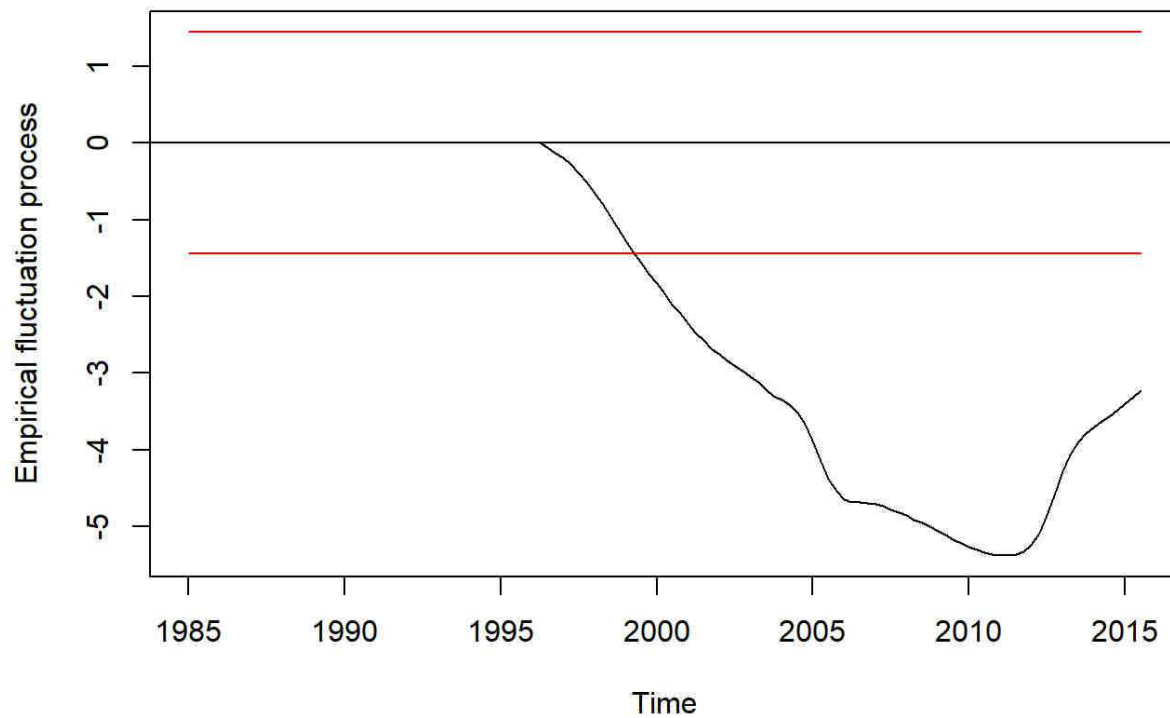


```
sctest(testestbr)
```

```
##  
## OLS-based CUSUM test  
##  
## data: testestbr  
## S0 = 5.3877, p-value < 2.2e-16
```

```
testestbr<-efp(modelo, type="Rec-MOSUM", data = datokunsim, h=0.2) # Este toma el error de predicci?  
n haciendo una Media Movil  
plot(testestbr)
```

## Recursive MOSUM test

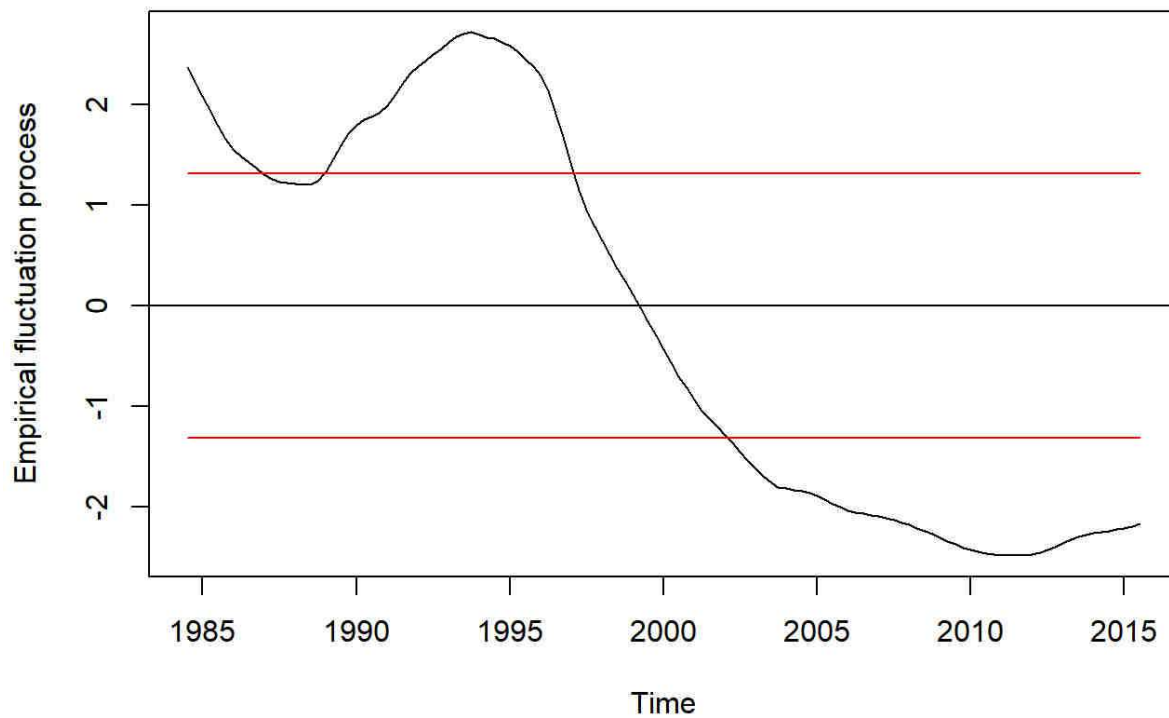


```
sctest(testestbr)
```

```
##  
## Recursive MOSUM test  
##  
## data: testestbr  
## M = 5.3814, p-value = 0.01
```

```
testestbr<-efp(modelo, type="OLS-MOSUM", data = datokunsim, h=0.2) # Este toma los residuos haciendo  
una Media Movil  
plot(testestbr)
```

## OLS-based MOSUM test



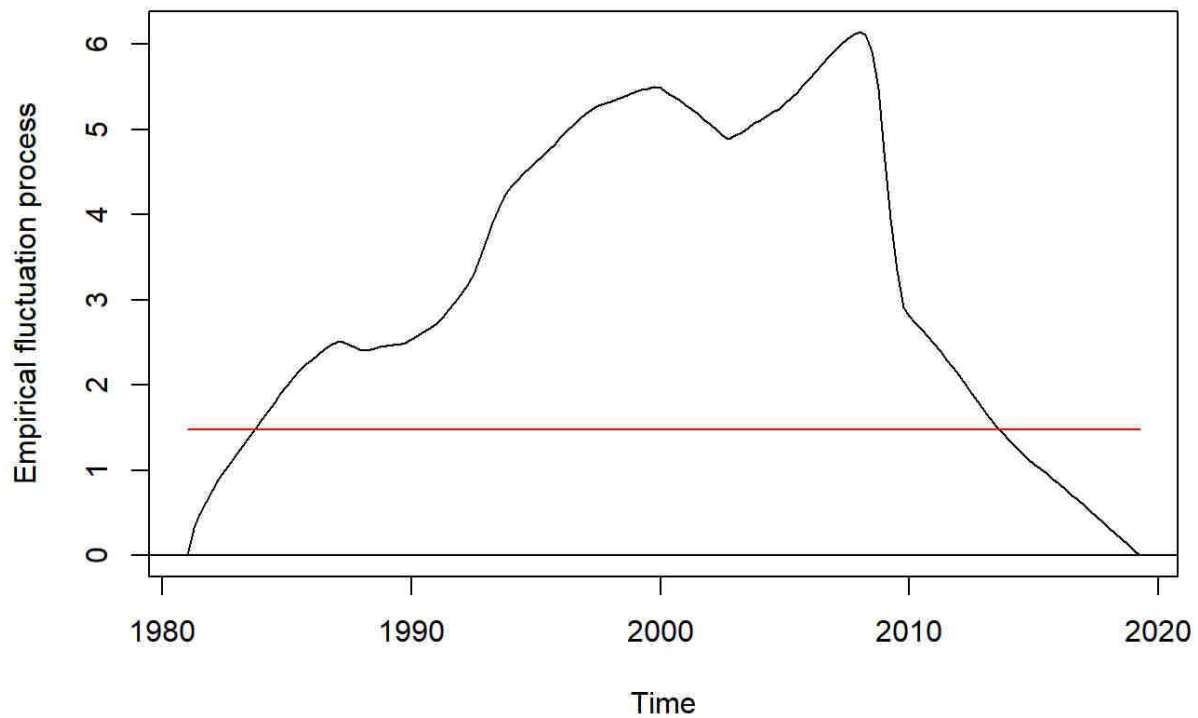
```
sctest(testestbr)
```

```
##  
## OLS-based MOSUM test  
##  
## data: testestbr  
## M0 = 2.7181, p-value = 0.01
```

```
#Para ver las diferencias entre los parametros
```

```
testestbr<-efp(modelo, type="RE", data = datokunsim, h=0.2) # este es para comprobar cambios en los  
par?metros Por MCrecurivos  
plot(testestbr)
```

## RE test (recursive estimates test)

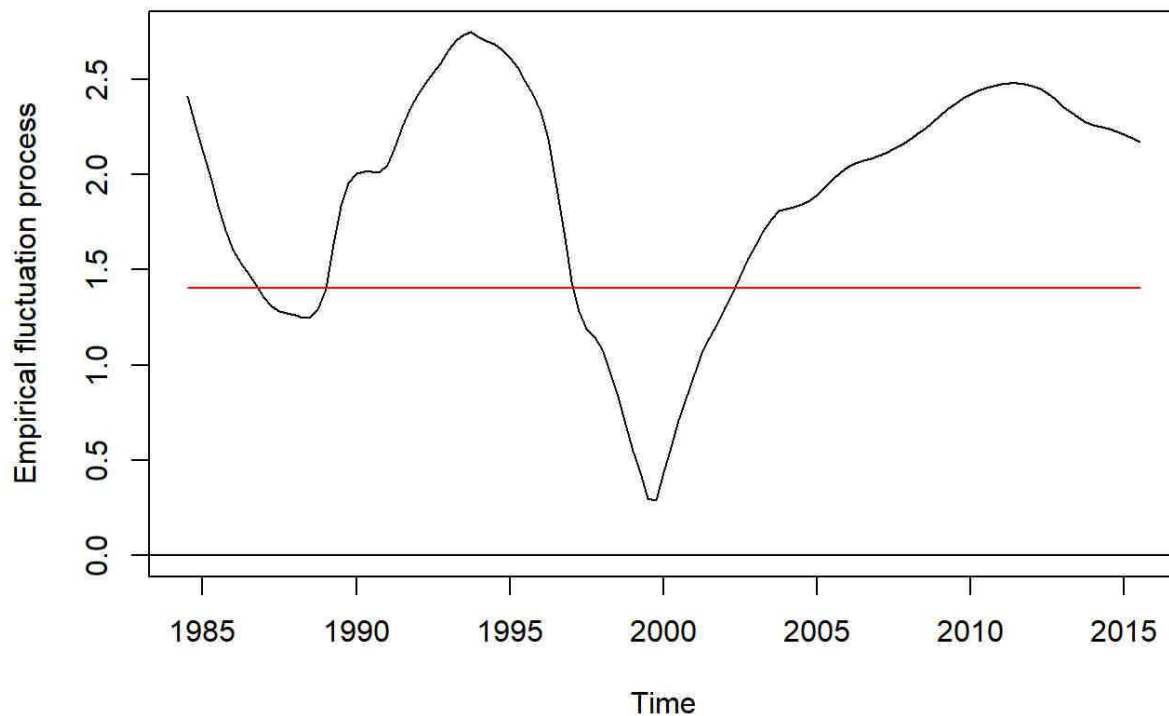


```
sctest(testestbr)
```

```
##  
## RE test (recursive estimates test)  
##  
## data: testestbr  
## RE = 6.1408, p-value < 2.2e-16
```

```
testestbr<-efp(modelo, type="ME", data = datokunsim, h=0.2) # este es para comprobar cambios en los  
parámetros Por Media Movel  
plot(testestbr)
```

## ME test (moving estimates test)



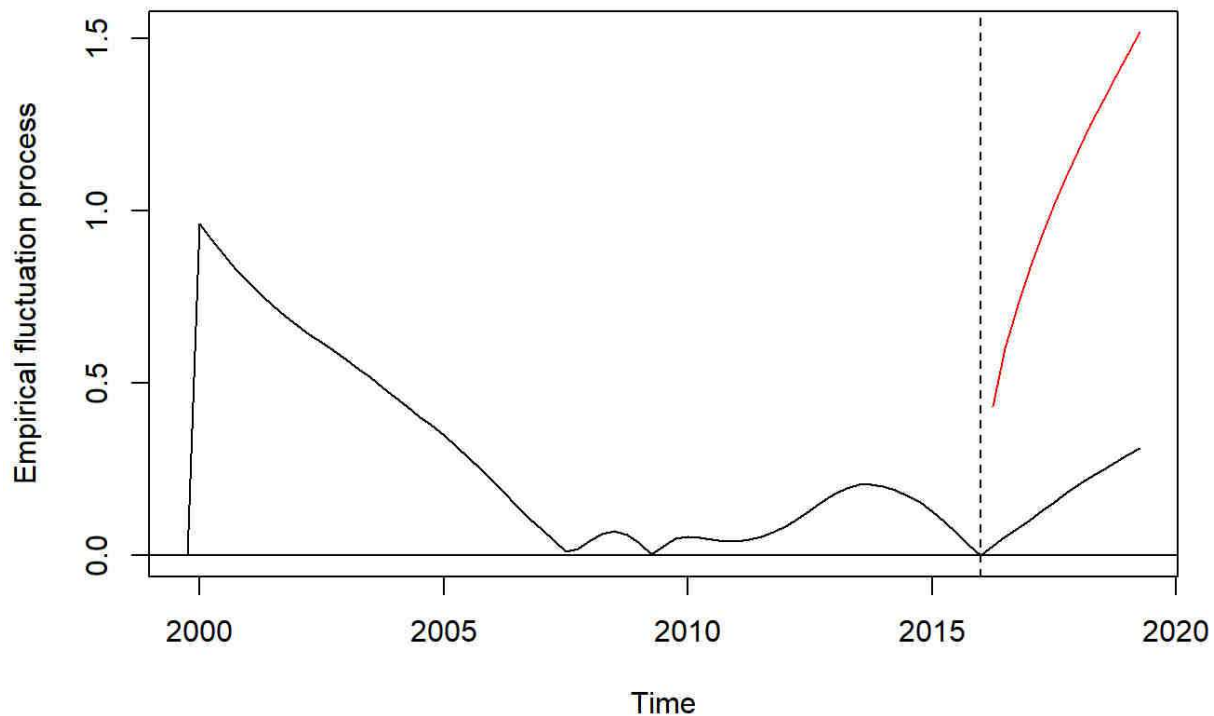
```
sctest(testestbr)
```

```
##  
## ME test (moving estimates test)  
##  
## data: testestbr  
## ME = 2.7464, p-value = 0.01
```

```
#  
# Los anteriores test sirven para saber si en las series históricas (en los datos pasados disponibles)  
# hay un cambio estructural. Cuando lo que se quiere saber es si los nuevos datos suponen un cambio  
# estructural puede compararse la regresión con los datos históricos con la regresión con los datos  
# históricos más los nuevos datos y analizar si hay diferencias significativas (Chow Forecast Test)  
  
# necesitamos crear un objeto "mefp"  
  
# Por ejemplo quiero ver si hay diferencias a partir del 2016  
datokunsim<-ts.union(T4ocu.ts=T4ocu_sim,T4pib.ts, dframe=FALSE)  
datokunsim2<-window(datokunsim,start=2000, end=2016, frequency=4)  
me.mefp <- mefp(modelo, type="OLS-CUSUM", data = datokunsim2, alpha = 0.05)  
  
# ahora añadimos los nuevos datos  
datokunsim2<-window(datokunsim,start=2000, end=tiempo_max, frequency=4)  
me.mefp <- monitor(me.mefp)  
  
plot(me.mefp)
```



## Monitoring with OLS-based CUSUM test



```
# No hay cambio estructural porqu no supera a la banda de significación (en rojo)
# Si hubiese un cambio estaría por encima del nivel de significación

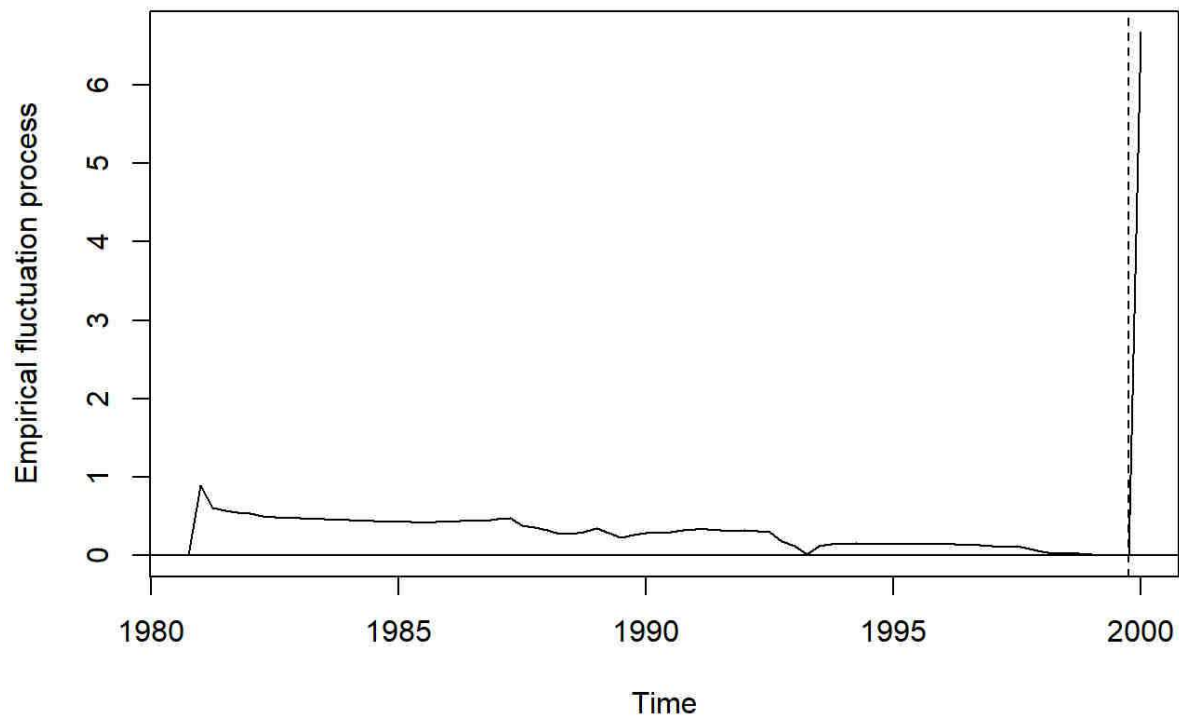
# Para que sea más evidente lo voy a hacer a partir del 2000
# sabemos que en el 2000 hay un cambio estructural (por construcción de las series)
# Supongamos que estamos en el 2001 ¿Cómo saber que hay un cambio estructural con sólo cuatro datos?
datokunsim<-ts.union(T4ocu.ts=T4ocu_sim,T4pib.ts, dframe=FALSE)
datokunsim2<-window(datokunsim,start=1981, end=1999.75, frequency=4)
me.mefp <- mefp(modelo, type="OLS-CUSUM", data = datokunsim2, alpha = 0.05)

# ahora añadimos los nuevos datos
datokunsim2<-window(datokunsim,start=1981, end=2000, frequency=4)
me.mefp <- monitor(me.mefp)
```

```
## Break detected at observation # 77
```

```
# !!!! ha detectado el cambio estructural del 2000
plot(me.mefp)
```

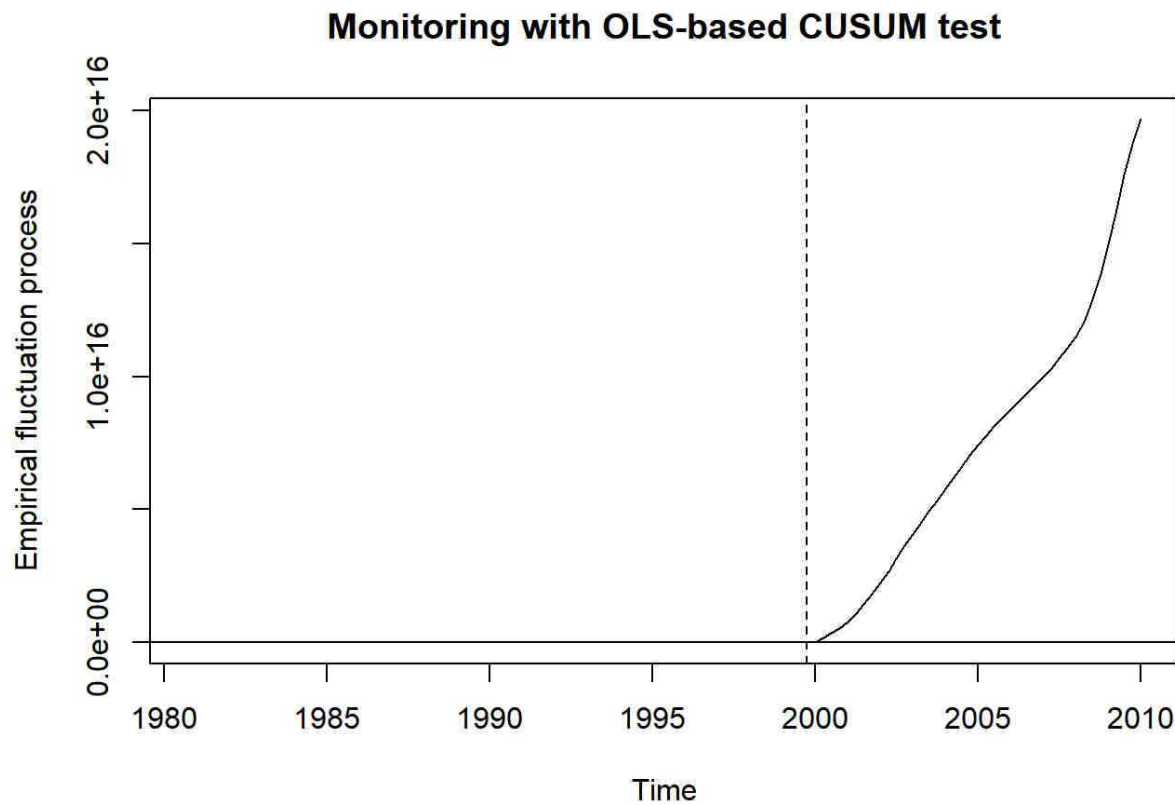
## Monitoring with OLS-based CUSUM test



```
datokunsim2<-window(datokunsim,start=1981, end=1999.75, frequency=4)
me.mefp <- mefp(modelo, type="OLS-CUSUM", data = datokunsim2, alpha = 0.05)
datokunsim2<-window(datokunsim,start=1981, end=2010, frequency=4)
me.mefp <- monitor(me.mefp)
```

```
## Break detected at observation # 77
```

```
plot(me.mefp)
```



4 Tarea punto 4\_ Cambio estructural: ¿Han sido eficaces las reformas que ha sufrido el mercado de trabajo?

```
#
#
# Tarea punto 4_ Cambio estructural

# Han sido eficaces las reformas que ha sufrido el mercado de trabajo?
# 1980 Estatuo de los trabajadores;
# 1984: inicio de la temporalidad;
# 1988: Nuevos contratos Temporales y basura;
# 1992: reforma del seguro desempleo;
# 1994: reforma importante que da entrada entre otras a la ETT;
# 1997: reforma pactada;
# 2001-2002: reforma Aznar con huelga y marcha atr?s;
# 2006: reforma leve; 2009: reforma leve;
# 2010 reforma importante ante la crisis continuada en 2012, con una Reforma importante

# Entenderemos "eficacia", como un cambio en la pendiente y en el punto de corte del
# eje x (con el mismo nivel de crecimiento econ?mico se puede crecer mas, y
# se puede emezas a crear empleo con un menor ritmo de crecimiento economico)

# Ahora comenzamos con la practica
# Divido en dos la muestra y hago dos regrsiones

datokun1<-ts.intersect(T4pib.ts,T4ocu.ts,dframe = FALSE)

# contraste 1984
datokun2<-window(datokun1,start=1981, end=1988.00, frequency=4)

modelo2.lm<-lm(modelo, data=datokun2, na.action=NULL)
summary(modelo2.lm)
```

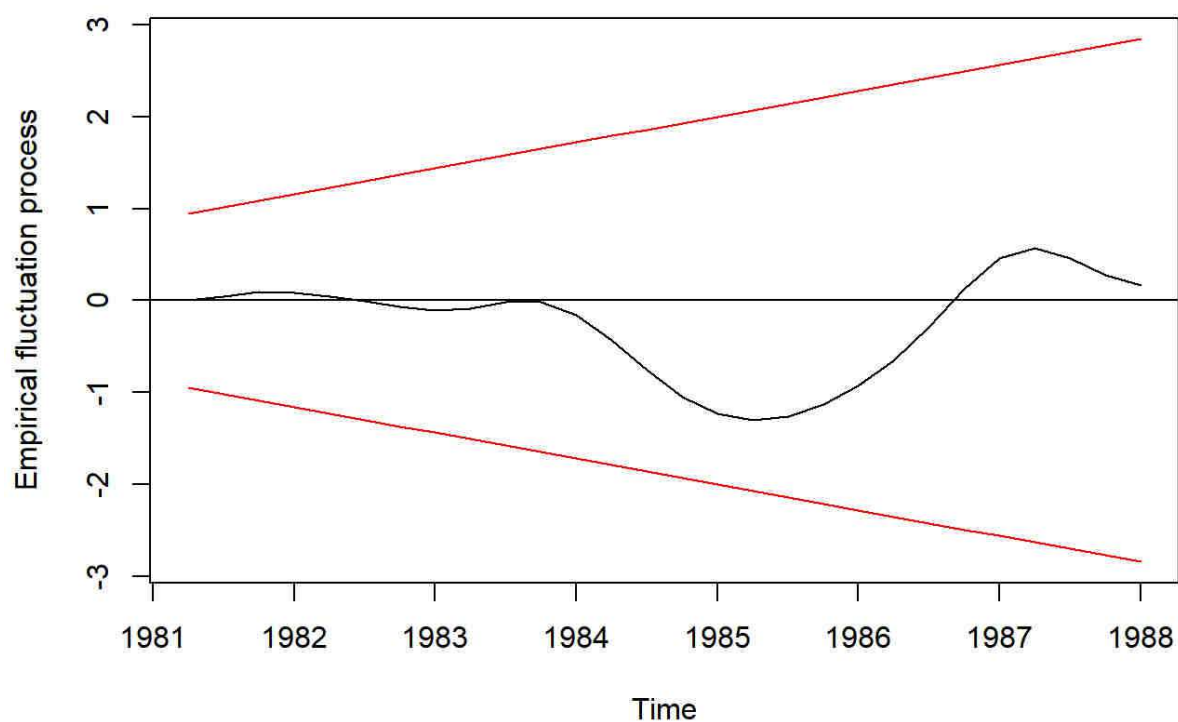
```
##
## Call:
## lm(formula = modelo, data = datokun2, na.action = NULL)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.022298 -0.006635  0.002566  0.006274  0.021389
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -0.029398   0.003543  -8.296 6.63e-09 ***
## T4pib.ts     1.285061   0.117931  10.897 2.18e-11 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.01157 on 27 degrees of freedom
## Multiple R-squared:  0.8147, Adjusted R-squared:  0.8079
## F-statistic: 118.7 on 1 and 27 DF,  p-value: 2.18e-11
```

```
stargazer(modelo2.lm, type="text")
```

```
##
## =====
##                      Dependent variable:
##                      -----
##                      T4ocu.ts
## -----
## T4pib.ts              1.285***
##                      (0.118)
##
## Constant              -0.029***
##                      (0.004)
## -----
## Observations          29
## R2                    0.815
## Adjusted R2           0.808
## Residual Std. Error   0.012 (df = 27)
## F Statistic           118.740*** (df = 1; 27)
## =====
## Note:                  *p<0.1; **p<0.05; ***p<0.01
```

```
testestbr<-efp(modelo, type="Rec-CUSUM", data = datokun2, h=0.2) # ESTE ES EL MISMO QUE EN EIEWS
plot(testestbr)
```

### Recursive CUSUM test



```
sctest(testestbr)
```

```
##
## Recursive CUSUM test
##
## data: testestbr
## S = 0.59551, p-value = 0.4221
```

```
#locator(10, type = "p" ,col=2, pch=19)
```

```
# contraste 1988
```

```
datokun3<-window(datokun1,start=1981, end=1992.75, frequency=4)
```

```
modelo3.lm<-lm(modelo, data=datokun3, na.action=NULL)
```

```
summary(modelo3.lm)
```

```
##
```

```
## Call:
```

```
## lm(formula = modelo, data = datokun3, na.action = NULL)
```

```
##
```

```
## Residuals:
```

```
##      Min       1Q   Median       3Q      Max
```

```
## -0.0247815 -0.0057071  0.0006119  0.0064003  0.0218639
```

```
##
```

```
## Coefficients:
```

```
##              Estimate Std. Error t value Pr(>|t|)
```

```
## (Intercept) -0.027409   0.002898  -9.459 2.32e-12 ***
```

```
## T4pib.ts     1.312553   0.088347  14.857 < 2e-16 ***
```

```
## ---
```

```
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
##
```

```
## Residual standard error: 0.01095 on 46 degrees of freedom
```

```
## Multiple R-squared:  0.8275, Adjusted R-squared:  0.8238
```

```
## F-statistic: 220.7 on 1 and 46 DF,  p-value: < 2.2e-16
```

```
stargazer(modelo2.lm, modelo3.lm, type="text")
```

```
##
```

```
## =====
```

```
##                               Dependent variable:
```

```
##                               -----
```

```
##                               T4ocu.ts
```

```
##                               (1)                (2)
```

```
## -----
```

```
## T4pib.ts                1.285***                1.313***
```

```
##                        (0.118)                (0.088)
```

```
##
```

```
## Constant                -0.029***                -0.027***
```

```
##                        (0.004)                (0.003)
```

```
##
```

```
## -----
```

```
## Observations                29                48
```

```
## R2                0.815                0.828
```

```
## Adjusted R2                0.808                0.824
```

```
## Residual Std. Error    0.012 (df = 27)    0.011 (df = 46)
```

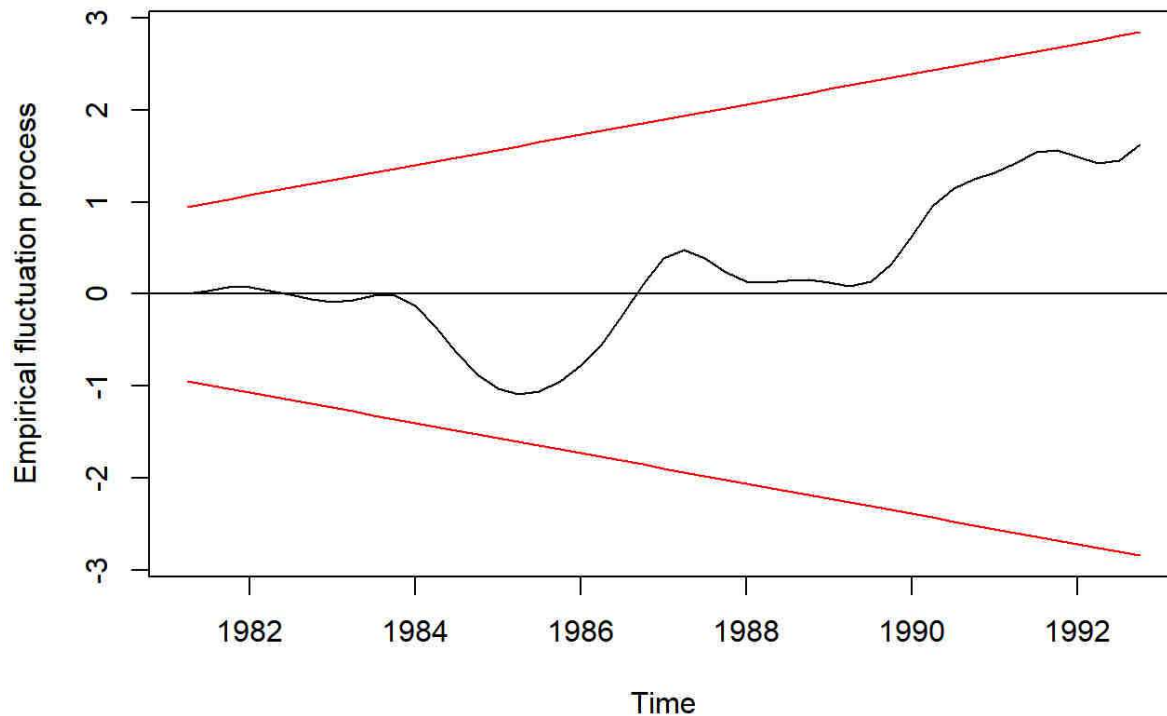
```
## F Statistic        118.740*** (df = 1; 27) 220.727*** (df = 1; 46)
```

```
## =====
```

```
## Note:                *p<0.1; **p<0.05; ***p<0.01
```

```
testestbr<-efp(modelo, type="Rec-CUSUM", data = datokun3, h=0.2) # ESTE ES EL MISMO QUE EN EVIEWS  
plot(testestbr)
```

## Recursive CUSUM test



```
#locator(10, type = "p" ,col=2, pch=19)
```

```
# Contrasto 1992
```

```
datokun4<-window(datokun1,start=1981, end=1994.75, frequency=4)
```

```
modelo4.lm<-lm(modelo, data=datokun4, na.action=NULL)
```

```
summary(modelo4.lm)
```

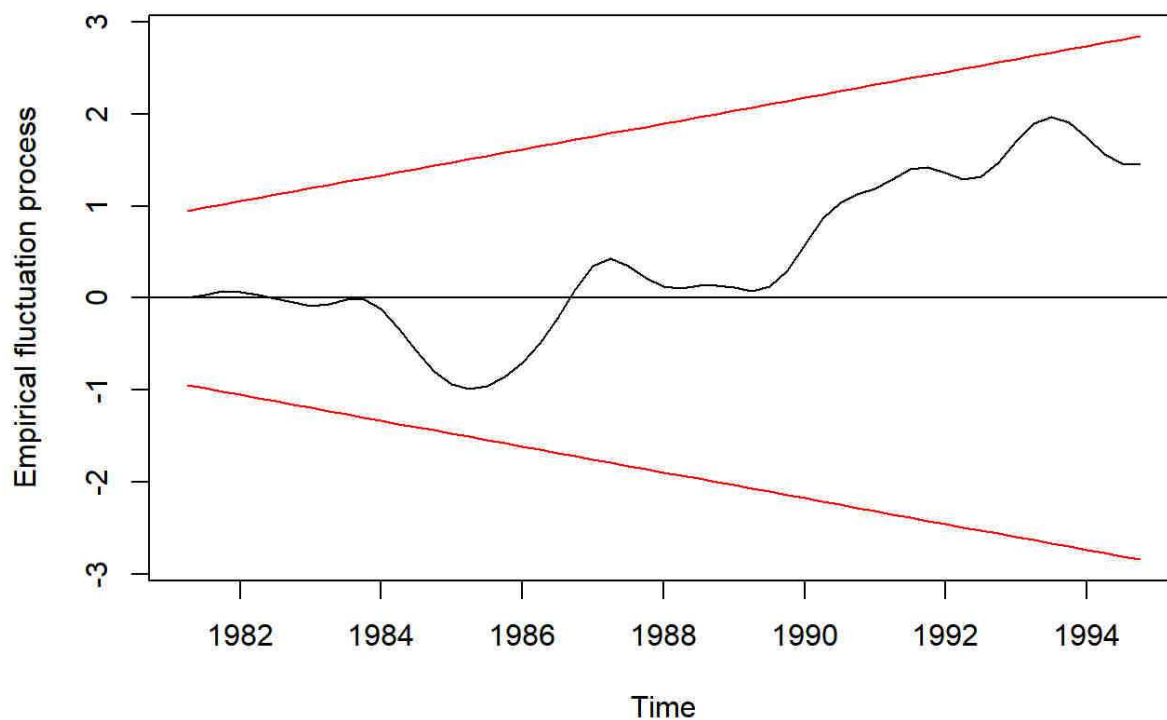
```
##
## Call:
## lm(formula = modelo, data = datokun4, na.action = NULL)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.0255805 -0.0060854 -0.0006126  0.0069210  0.0226990
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -0.025011   0.002384  -10.49 1.23e-14 ***
## T4pib.ts     1.223468   0.076264   16.04 < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.01106 on 54 degrees of freedom
## Multiple R-squared:  0.8266, Adjusted R-squared:  0.8234
## F-statistic: 257.4 on 1 and 54 DF, p-value: < 2.2e-16
```

```
stargazer(modelo2.lm, modelo3.lm,modelo4.lm, type="text")
```

```
##
## =====
##                               Dependent variable:
##                               -----
##                               T4ocu.ts
##                               (1)          (2)          (3)
## -----
## T4pib.ts          1.285***          1.313***          1.223***
##                   (0.118)          (0.088)          (0.076)
##
## Constant          -0.029***          -0.027***          -0.025***
##                   (0.004)          (0.003)          (0.002)
## -----
## Observations              29              48              56
## R2                        0.815            0.828            0.827
## Adjusted R2              0.808            0.824            0.823
## Residual Std. Error    0.012 (df = 27)    0.011 (df = 46)    0.011 (df = 54)
## F Statistic            118.740*** (df = 1; 27) 220.727*** (df = 1; 46) 257.363*** (df = 1; 54)
## =====
## Note:                                     *p<0.1; **p<0.05; ***p<0.01
```

```
testestbr<-efp(modelo, type="Rec-CUSUM", data = datokun4, h=0.2) # ESTE ES EL MISMO QUE EN EVIEWS
plot(testestbr)
```

### Recursive CUSUM test





```
#locator(10, type = "p" ,col=2, pch=19)
```

```
# Contrasto 1994
```

```
datokun5<-window(datokun1,start=1981, end=1997.75, frequency=4)
```

```
modelo5.lm<-lm(modelo, data=datokun5, na.action=NULL)
```

```
summary(modelo5.lm)
```

```
##
```

```
## Call:
```

```
## lm(formula = modelo, data = datokun5, na.action = NULL)
```

```
##
```

```
## Residuals:
```

```
##      Min       1Q   Median       3Q      Max  
## -0.0269559 -0.0057394 -0.0002265  0.0061677  0.0207315
```

```
##
```

```
## Coefficients:
```

```
##           Estimate Std. Error t value Pr(>|t|)
```

```
## (Intercept) -0.02422    0.00233  -10.39 1.56e-15 ***
```

```
## T4pib.ts     1.25575    0.07452   16.85 < 2e-16 ***
```

```
## ---
```

```
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
##
```

```
## Residual standard error: 0.011 on 66 degrees of freedom
```

```
## Multiple R-squared:  0.8114, Adjusted R-squared:  0.8085
```

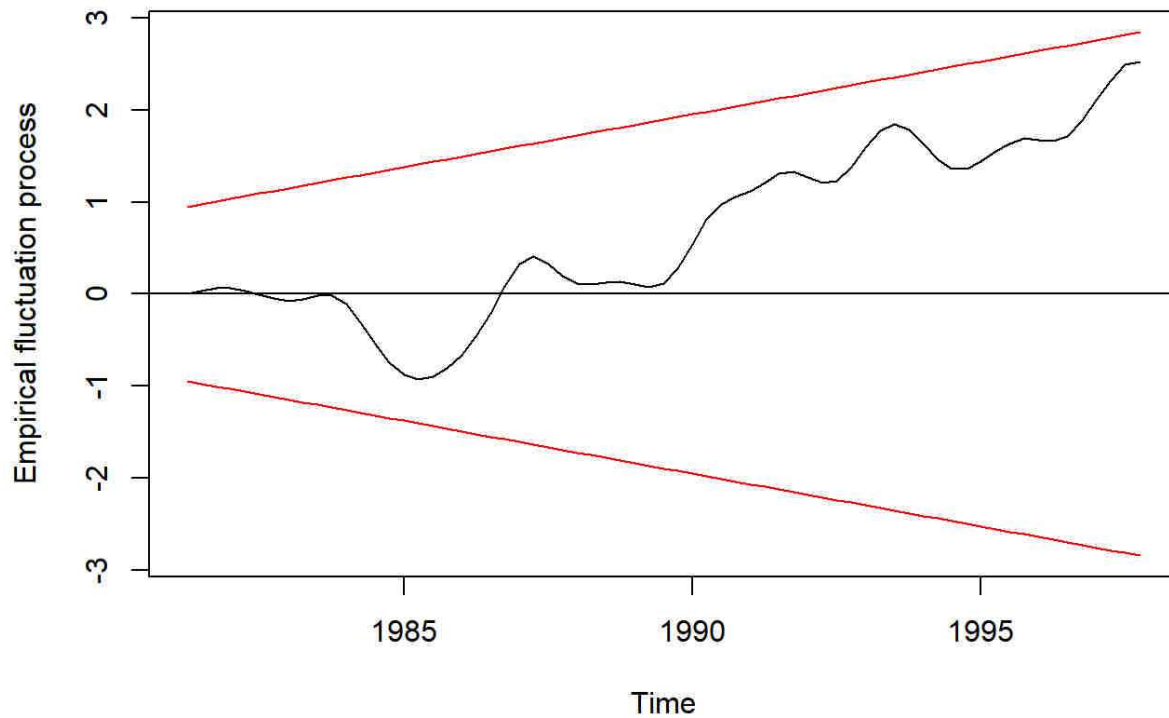
```
## F-statistic: 284 on 1 and 66 DF, p-value: < 2.2e-16
```

```
stargazer(modelo2.lm, modelo3.lm,modelo4.lm,modelo5.lm, type="text")
```

```
##
## =====
=====
##                               Dependent variable:
##                               -----
##                               T4ocu.ts
##                               (1)           (2)           (3)
(4)
## -----
## T4pib.ts           1.285***           1.313***           1.223***
1.256***
##           (0.118)           (0.088)           (0.076)
(0.075)
##
## Constant           -0.029***           -0.027***           -0.025***
-0.024***
##           (0.004)           (0.003)           (0.002)
(0.002)
##
## -----
## Observations           29           48           56
68
## R2           0.815           0.828           0.827
0.811
## Adjusted R2           0.808           0.824           0.823
0.809
## Residual Std. Error   0.012 (df = 27)   0.011 (df = 46)   0.011 (df = 54)
0.011 (df = 66)
## F Statistic       118.740*** (df = 1; 27) 220.727*** (df = 1; 46) 257.363*** (df = 1; 54) 283.9
58*** (df = 1; 66)
## =====
=====
## Note:
##                               *p<0.1; *
*p<0.05; ***p<0.01
```

```
testestbr<-efp(modelo, type="Rec-CUSUM", data = datokun5, h=0.2) # ESTE ES EL MISMO QUE EN EVIEWS
plot(testestbr)
```

## Recursive CUSUM test



```
#locator(10, type = "p" ,col=2, pch=19)
```

```
# Contrasto 1997
```

```
datokun6<-window(datokun1,start=1981, end=2002.75, frequency=4)
```

```
modelo6.lm<-lm(modelo, data=datokun6, na.action=NULL)
```

```
summary(modelo6.lm)
```

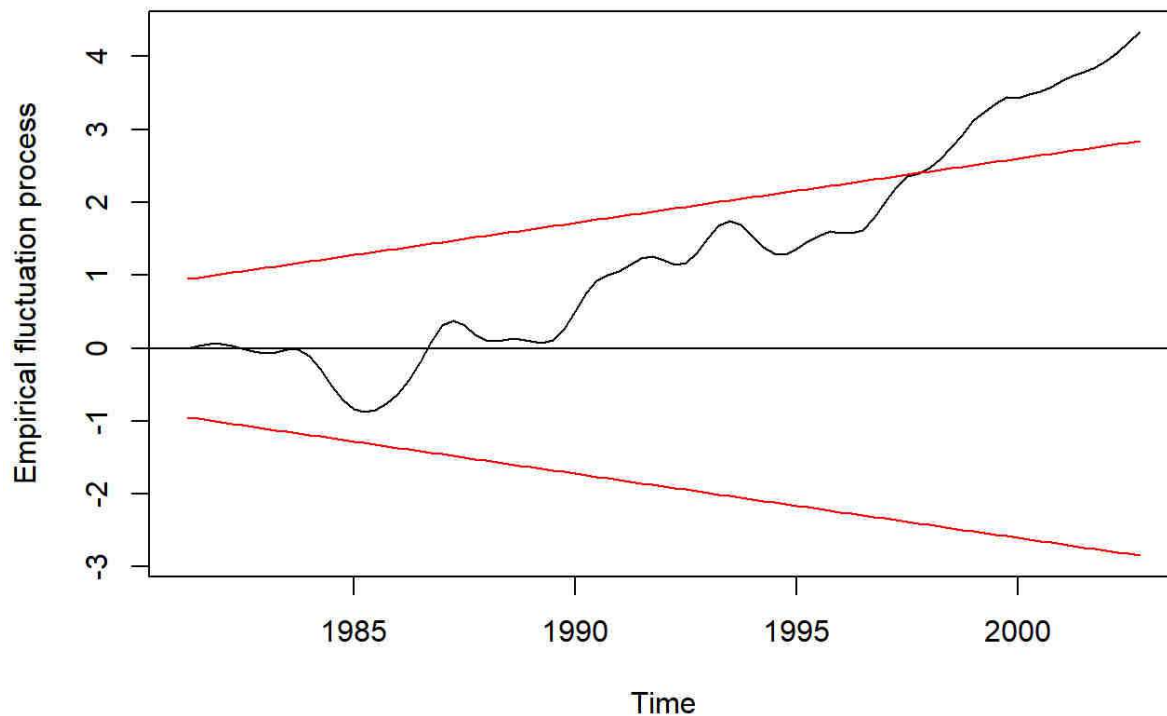
```
##
## Call:
## lm(formula = modelo, data = datokun6, na.action = NULL)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.028437 -0.006520  0.001971  0.007500  0.017620
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -0.024330   0.002227  -10.92  <2e-16 ***
## T4pib.ts     1.344634   0.065113   20.65  <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.01075 on 86 degrees of freedom
## Multiple R-squared:  0.8322, Adjusted R-squared:  0.8302
## F-statistic: 426.5 on 1 and 86 DF, p-value: < 2.2e-16
```

```
stargazer(modelo2.lm, modelo3.lm,modelo4.lm,modelo5.lm,modelo6.lm, type="text")
```

```
##
## =====
=====
##                                     Dependent variable:
## -----
##                                     T4ocu.ts
##                                     (3)
##          (1)          (2)
(4)          (5)
## -----
## T4pib.ts          1.285***          1.313***          1.223***
1.256***          1.345***
##          (0.118)          (0.088)          (0.076)
(0.075)          (0.065)
##
## Constant          -0.029***          -0.027***          -0.025***
-0.024***          -0.024***
##          (0.004)          (0.003)          (0.002)
(0.002)          (0.002)
##
## -----
## Observations          29          48          56
68          88
## R2          0.815          0.828          0.827
0.811          0.832
## Adjusted R2          0.808          0.824          0.823
0.809          0.830
## Residual Std. Error    0.012 (df = 27)    0.011 (df = 46)    0.011 (df = 54)
0.011 (df = 66)    0.011 (df = 86)
## F Statistic    118.740*** (df = 1; 27) 220.727*** (df = 1; 46) 257.363*** (df = 1; 54) 283.9
58*** (df = 1; 66) 426.453*** (df = 1; 86)
## =====
=====
## Note:
*p<0.1; **p<0.05; ***p<0.01
```

```
testestbr<-efp(modelo, type="Rec-CUSUM", data = datokun6, h=0.2) # ESTE ES EL MISMO QUE EN EVIEWS
plot(testestbr)
```

## Recursive CUSUM test



```
#locator(10, type = "p" ,col=2, pch=19)
```

```
# primera regresion
datokun97<-window(datokun1,start=1981, end=1997.75, frequency=4)
modelo97.lm<-lm(modelo, data=datokun97, na.action=NULL)
summary(modelo97.lm)
```

```
##
## Call:
## lm(formula = modelo, data = datokun97, na.action = NULL)
##
## Residuals:
```

	Min	1Q	Median	3Q	Max
	-0.0269559	-0.0057394	-0.0002265	0.0061677	0.0207315

```
##
## Coefficients:
```

	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	-0.02422	0.00233	-10.39	1.56e-15 ***
T4pib.ts	1.25575	0.07452	16.85	< 2e-16 ***

```
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.011 on 66 degrees of freedom
## Multiple R-squared:  0.8114, Adjusted R-squared:  0.8085
## F-statistic: 284 on 1 and 66 DF, p-value: < 2.2e-16
```

```
stargazer(modelo.lm, modelo97.lm, type="text")
```

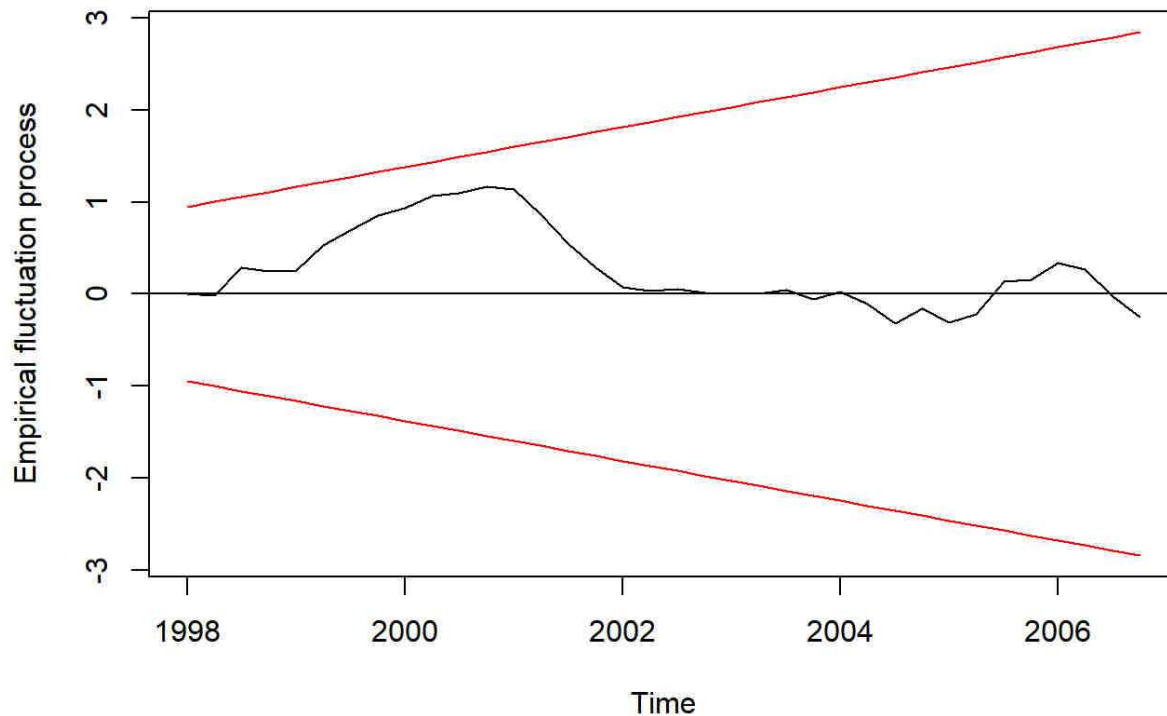
```
##
## =====
##                               Dependent variable:
##                               -----
##                               T4ocu.ts
##                               (1)                (2)
## -----
## T4pib.ts                1.244***                1.256***
##                        (0.037)                (0.075)
##
## Constant                -0.017***                -0.024***
##                        (0.001)                (0.002)
## -----
## Observations                154                68
## R2                        0.884                0.811
## Adjusted R2                0.883                0.809
## Residual Std. Error      0.010 (df = 152)      0.011 (df = 66)
## F Statistic              1,159.712*** (df = 1; 152) 283.958*** (df = 1; 66)
## =====
## Note:                      *p<0.1; **p<0.05; ***p<0.01
```

```
# Contrasto 2002
datokun7<-window(datokun1,start=1997.75, end=2006.75, frequency=4)
modelo7.lm<-lm(modelo, data=datokun7, na.action=NULL)
summary(modelo7.lm)
```

```
##
## Call:
## lm(formula = modelo, data = datokun7, na.action = NULL)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.0063915 -0.0028065  0.0000955  0.0017970  0.0092325
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -0.002079   0.003270  -0.636   0.529
## T4pib.ts     0.976639   0.081716  11.952 6.56e-14 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.003725 on 35 degrees of freedom
## Multiple R-squared:  0.8032, Adjusted R-squared:  0.7976
## F-statistic: 142.8 on 1 and 35 DF, p-value: 6.565e-14
```

```
testestbr<-efp(modelo, type="Rec-CUSUM", data = datokun7, h=0.2) # ESTE ES EL MISMO QUE EN EVIEWS
plot(testestbr)
```

## Recursive CUSUM test



```
# locator(10, type = "p" ,col=2, pch=19)
```

## 5 Ahora Vosotros

5.1 Contrastad si ha habido cambio en 2006

5.2 Contrastad si ha habido cambio en 2009

5.3 Contrastad si ha habido cambio 2010

5.4 Contrastad si ha habido cambio 2012

Deberían salir dos cambios estructurales, uno tras la reforma de 1997 y otro tras la última reforma de 2010-2012  
¿cuantas regresiones diferentes existen entonces? ¿Podrías realizar el test de Chow para los años en los que el tes  
REC-CUSUM haya sido significativo? Tendrías que volver a responder a la pregunta 2 y 3 de la práctica a la vista de  
los resultados de estos test de cambio estructural?

```
#####33  
#####33
```



Proyecto de innovación  
“Innova Docencia” de la UCM  
Convocatoria 2018-2019  
Proyecto número 275

**R-adaptación de la asignatura de métodos econométricos en economía y finanzas del  
Grado de Estadística Aplicada**

Facultad de Estudios Estadísticos  
Departamento de Economía Aplicada, Pública y Política

**Título**

***Práctica 4 : Ley de Okun con datos de la UE desde la web de Eurostat***

**Correspondiente al**

**Capítulo 2 REGRESIÓN LINEAL**

- El Modelo Lineal General (MLG)
- Hipótesis del modelo
- Estimación por Mínimos Cuadrados Ordinarios (MCO)
- Inferencia y predicción en el MLG
- Contrastes de cambio estructural
- Multicolinealidad y errores de medida

***Práctica 2 : La Ley de Okun: ¿a qué ritmo debe crecer la Economía para comenzar a generar empleo***

***Práctica 3 : La Ley de Okun: Test de Cambio Estructural, ¿han sido “eficaces” las reformas del mercado de trabajo?***

***Práctica 4 : Ley de Okun con datos de la UE desde la web de Eurostat***

***Práctica 5 : La ecuación de salarios y la discriminación por razón de género***



# Ley de Okun con datos de Eurostat

## *Asignatura* *Métodos Econométricos en Economía y Finanzas*

*Profesores*  
*Loren Escot, Alicia Pérez-Alonso y Julio E. Sandubete*

Dpto Economía Aplicada, Pública y Política  
Facultad Estudios Estadísticos - Universidad Complutense de Madrid

*Curso académico 2018/2019*

### **Abstract**

En esta práctica se muestra cómo utilizar la librería `eurostat` para descargar datos directamente de eurstat, pintar mapas, y gráficos. Se utilizaran datos de la unión europea para analizar qué países están haciéndolo relativamente mejor en términos de creación de empleo

- 1 Librerías necesarias
- 2 Búsqueda de datos en las tablas de eurostat
- 3 Representación de la importancia de cada país en el conjunto de la EU
- 4 Tasas de Crecimiento del PIB
- 5 Tasas de Crecimiento de los Ocupados
- 6 Ley de Okun en la Unión Europea

## 1 Librerías necesarias

```
library(eurostat)
```

```
## Warning: package 'eurostat' was built under R version 3.4.3
```

```
library(dplyr)
```

```
## Warning: package 'dplyr' was built under R version 3.4.4
```

```
##  
## Attaching package: 'dplyr'
```

```
## The following objects are masked from 'package:stats':  
##  
##   filter, lag
```

```
## The following objects are masked from 'package:base':  
##  
##   intersect, setdiff, setequal, union
```

```
library(ggplot2)
```

```
## Warning: package 'ggplot2' was built under R version 3.4.4
```

```
library(stargazer)
```

```
## Warning: package 'stargazer' was built under R version 3.4.1
```

```
##  
## Please cite as:
```

```
## Hlavac, Marek (2015). stargazer: Well-Formatted Regression and Summary Statistics Tables.
```

```
## R package version 5.2. http://CRAN.R-project.org/package=stargazer
```

## 2 Búsqueda de datos en las tablas de eurostat

```
# Busco datos de PIB en el conjunto de tablas de Eurostat  
query<-search_eurostat("GDP", type="table")  
  
query[1:10,1:2]
```

```
## # A tibble: 10 x 2  
##   title                                code  
##   <chr>                               <chr>  
## 1 GDP per capita in PPS                tec00114  
## 2 Real GDP growth rate - volume        tec00115  
## 3 Exports of goods and services in % of GDP tet00003  
## 4 Imports of goods and services in % of GDP tet00004  
## 5 Real GDP per capita                  sdg_08_10  
## 6 GDP deflator                         teina110  
## 7 GDP per capita in PPS                tec00114  
## 8 Volume of passenger transport relative to GDP ttr00001  
## 9 Percentage of the ICT sector on GDP    tin00074  
## 10 Private sector credit flow, consolidated - % GDP tipspc20
```

revisando las posibles tablas encontramos una titulada “GDP and main components (output, expenditure and income)” utilizaremos entonces esta tabla que se llama “nama\_10\_gdp”

```
dat.pib<-get_eurostat("nama_10_gdp",time_format = "num",type ="code", stringsAsFactors = TRUE) #,time_format = "raw",type ="code", stringsAsFactors = TRUE)
```

```
## Table nama_10_gdp cached at C:\Users\user\AppData\Local\Temp\Rtmpk3Zcbw/eurostat/nama_10_gdp_num_code_TF.rds
```

```
names(dat.pib)
```

```
## [1] "unit"      "na_item"   "geo"       "time"      "values"
```

```
#Para ver que significan los campos  
cc<-label_eurostat_vars(dat.pib)  
print(cc)
```

```
## [1] "Unit of measure"
## [2] "National accounts indicator (ESA 2010)"
## [3] "Geopolitical entity (reporting)"
## [4] "Period of time (a=annual, q=quarterly, m=monthly, d=daily, c=cumulated from January)"
```

### 3 Representación de la importancia de cada país en el conjunto de la EU

```
table(dat.pib$time)
```

```
##
## 1975 1976 1977 1978 1979 1980 1981 1982 1983 1984 1985 1986
## 1394 1447 1447 1447 1447 2030 2187 2193 2193 2193 2193 2193
## 1987 1988 1989 1990 1991 1992 1993 1994 1995 1996 1997 1998
## 2193 2193 2193 2266 2630 2761 3076 3217 18291 21667 22164 22459
## 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 22588 24614 24972 25097 25153 25156 25156 25366 25534 26296 26325 26325
## 2011 2012 2013 2014 2015 2016 2017 2018
## 26325 26325 26386 26386 26382 26363 25958 20658
```

```
dat.pib<-dat.pib%>%
  filter(time==2017 | time==2018)

# añadir los labels
dat.pib<-label_eurostat(dat.pib, code=names(dat.pib))

#View(dat.pib)

table(dat.pib$na_item)
```

##	
##	Value added, gross
##	1809
##	Gross domestic product at market prices
##	1888
##	Taxes on products
##	1132
##	Taxes less subsidies on products
##	1809
##	Subsidies on products
##	1132
##	Final consumption expenditure
##	1858
##	Individual consumption expenditure of general government
##	1757
##	Final consumption expenditure of households
##	1794
##	Household and NPISH final consumption expenditure
##	1858
##	Final consumption expenditure of NPISH
##	1757
##	Collective consumption expenditure of general government
##	1757
##	Final consumption expenditure and gross capital formation
##	1847
##	Final consumption expenditure, gross capital formation and exports of goods and services
##	1847
##	Final consumption expenditure of general government
##	1871
##	Actual individual consumption
##	1770
##	Gross fixed capital formation
##	1827
##	Gross capital formation
##	1858
##	Exports of goods and services
##	1871
##	Exports of goods
##	1814
##	Exports of services
##	1814
##	Imports of goods and services
##	1871
##	Imports of goods
##	1814
##	Imports of services
##	1814
##	Changes in inventories and acquisitions less disposals of valuables
##	586
##	External balance of goods and services
##	336
##	External balance - Goods
##	330
##	External balance - Services
##	330
##	Operating surplus and mixed income, gross
##	538
##	Compensation of employees
##	538
##	Wages and salaries
##	538

##	Employers' social contributions	
##		538
##	Taxes on production and imports	
##		253
##	Taxes on production and imports less subsidies	
##		538
##	Subsidies	
##		253
##	Changes in inventories	
##		339
##	Acquisitions less disposals of valuables	
##		296
##	Statistical discrepancy (expenditure approach)	
##		238
##	Statistical discrepancy (production approach)	
##		255
##	Statistical discrepancy (income approach)	
##		141

```

dat.pib<-dat.pib%>%
  filter(na_item=="Gross domestic product at market prices")

table(dat.pib$unit)

```

##		
##	Chain linked volumes (2005), million eur	
o		
##		8
3		
##	Chain linked volumes (2005), million units of national currenc	
y		
##		8
3		
##	Chain linked volumes (2010), million eur	
o		
##		8
5		
##	Chain linked volumes (2010), million units of national currenc	
y		
##		8
5		
##	Chain linked volumes, index 2005=10	
0		
##		8
3		
##	Chain linked volumes, index 2010=10	
0		
##		8
5		
##	Chain linked volumes, percentage change on previous perio	
d		
##		8
5		
##	Contribution to GDP growth, percentage point change on previous perio	
d		
##		1
1		
##	Current prices, million eur	
o		
##		8
6		
##	Current prices, million units of national currenc	
y		
##		8
6		
##	Current prices, million purchasing power standard	
s		
##		5
2		
##	Percentage of EU27 (from 2019) total (based on million euro), current price	
s		
##		8
6		
##	Percentage of EU28 total (based on million euro), current price	
s		
##		8
6		
##	Percentage of EU28 total (based on million purchasing power standards), current price	
s		
##		5
2		
##	Percentage of gross domestic product (GD	
P)		
##		8
6		

```

## Price index (implicit deflator), 2005=100, eur
0
## 8
3
## Price index (implicit deflator), 2005=100, national currenc
y
## 8
3
## Price index (implicit deflator), 2010=100, eur
0
## 8
5
## Price index (implicit deflator), 2010=100, national currenc
y
## 8
5
## Price index (implicit deflator), percentage change on previous period, eur
0
## 8
3
## Price index (implicit deflator), percentage change on previous period, national currenc
y
## 8
3
## Previous year prices, million eur
0
## 8
4
## Previous year prices, million units of national currenc
y
## 8
2
## Current prices, million purchasing power standards (PPS, EU27 from 201
9)
## 4
3
## Percentage of EU27 (from 2019) total (based on million purchasing power standards), current price
s
## 4
3

```

```

# Voy a hacer un mapa y un gráfico para analizar el peso de cada país

```

```

dat.pib_mapa<-dat.pib%>%
  filter(unit_code=="PC_EU28_MPPS_CP")%>%
  filter(time==2017)

```

```

# Cambio algunos nombres que están muy largos
levels(dat.pib_mapa$geo)

```

```
## [1] "Albania"
## [2] "Austria"
## [3] "Bosnia and Herzegovina"
## [4] "Belgium"
## [5] "Bulgaria"
## [6] "Switzerland"
## [7] "Cyprus"
## [8] "Czechia"
## [9] "Germany (until 1990 former territory of the FRG)"
## [10] "Denmark"
## [11] "Euro area (EA11-2000, EA12-2006, EA13-2007, EA15-2008, EA16-2010, EA17-2013, EA18-2014, EA19)"
## [12] "Euro area (12 countries)"
## [13] "Euro area (19 countries)"
## [14] "Estonia"
## [15] "Greece"
## [16] "Spain"
## [17] "European Union - 15 countries (1995-2004)"
## [18] "European Union - 27 countries (from 2019)"
## [19] "European Union - 28 countries"
## [20] "Finland"
## [21] "France"
## [22] "Croatia"
## [23] "Hungary"
## [24] "Ireland"
## [25] "Italy"
## [26] "Lithuania"
## [27] "Luxembourg"
## [28] "Latvia"
## [29] "North Macedonia"
## [30] "Netherlands"
## [31] "Norway"
## [32] "Poland"
## [33] "Portugal"
## [34] "Romania"
## [35] "Serbia"
## [36] "Sweden"
## [37] "Slovenia"
## [38] "Slovakia"
## [39] "United Kingdom"
## [40] "Iceland"
## [41] "Malta"
## [42] "Turkey"
## [43] "Montenegro"
## [44] "Kosovo (under United Nations Security Council Resolution 1244/99)"
## [45] "Liechtenstein"
```

```
levels(dat.pib_mapa$geo)[9]<-"Germany"
levels(dat.pib_mapa$geo)[44]<-"Kosovo"
levels(dat.pib_mapa$geo)[17]<-"15 EU"
levels(dat.pib_mapa$geo)[18]<-"27 EU"
levels(dat.pib_mapa$geo)[19]<-"28 EU"
levels(dat.pib_mapa$geo)[11]<-"EURO AREA"
levels(dat.pib_mapa$geo)[12]<-"12-EURO AREA"
levels(dat.pib_mapa$geo)[13]<-"19-EURO AREA"

# Elimino algunos datos de agregados EU y EUROAREA
levels(dat.pib_mapa$geo)
```



```
## [1] "Albania" "Austria"
## [3] "Bosnia and Herzegovina" "Belgium"
## [5] "Bulgaria" "Switzerland"
## [7] "Cyprus" "Czechia"
## [9] "Germany" "Denmark"

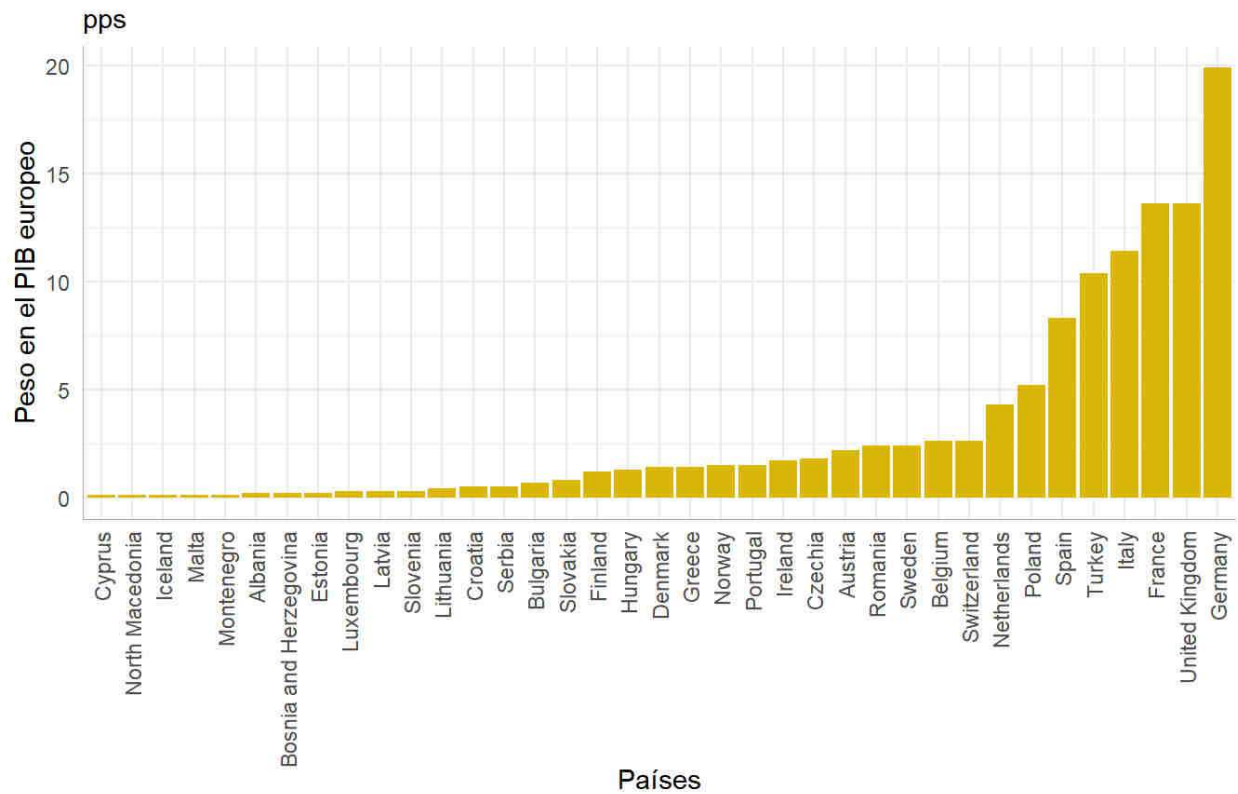
## [11] "EURO AREA" "12-EURO AREA"
## [13] "19-EURO AREA" "Estonia"
## [15] "Greece" "Spain"
## [17] "15 EU" "27 EU"
## [19] "28 EU" "Finland"
## [21] "France" "Croatia"
## [23] "Hungary" "Ireland"
## [25] "Italy" "Lithuania"
## [27] "Luxembourg" "Latvia"
## [29] "North Macedonia" "Netherlands"
## [31] "Norway" "Poland"
## [33] "Portugal" "Romania"
## [35] "Serbia" "Sweden"
## [37] "Slovenia" "Slovakia"
## [39] "United Kingdom" "Iceland"
## [41] "Malta" "Turkey"
## [43] "Montenegro" "Kosovo"
## [45] "Liechtenstein"
```

```
dat.pib_mapa<-dat.pib_mapa[c(-11,-12,-13,-17,-18,-19),]
```

```
g1<-ggplot() +
  geom_bar(data=dat.pib_mapa, aes(x=reorder(geo, values),y=values) ,fill = "#D8B70A",stat = "identity")+
  ylab("Peso en el PIB europeo")+
  xlab("Países")+
  labs( title="Participación en el PIB europeo (%)",
        subtitle="pps",
        caption = paste("Fuente: Eurostat", sep = "\n"))+
  theme_minimal()+
  theme(legend.position = "right",
        plot.title= element_text(colour="black", face="bold"),
        axis.line = element_line(size = 0.25, colour = "darkgrey", linetype = 1),
        legend.title=element_blank(),
        axis.text.x = element_text(angle = 90, hjust=1, vjust = 0.5))

print(g1)
```

## Participación en el PIB europeo (%)



Fuente: Eurostat

```
#Mapa
aamap<-dat.pib_mapa%>%
  mutate(cat = cut_to_classes(values, n=7,style="quantile", decimals=1))%>%
  mutate(geo = geo_code)

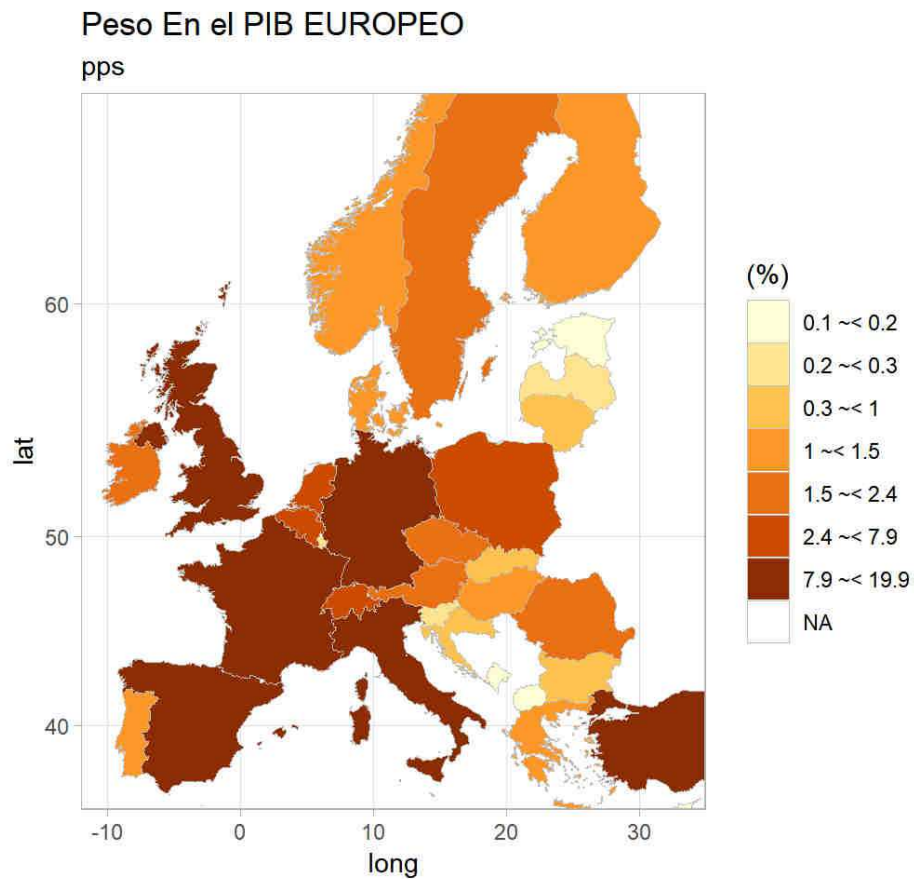
mapdata <- merge_eurostat_geodata(aamap, all_regions = TRUE,
                                   resolution = "10")
```

```
##
##      COPYRIGHT NOTICE
##
##      When data downloaded from this page
##      <http://ec.europa.eu/eurostat/web/gisco/geodata/reference-data/administrative-units-statistical-units>
##      is used in any printed or electronic publication,
##      in addition to any other provisions
##      applicable to the whole Eurostat website,
##      data source will have to be acknowledged
##      in the legend of the map and
##      in the introductory page of the publication
##      with the following copyright notice:
##
##      - EN: (C) EuroGeographics for the administrative boundaries
##      - FR: (C) EuroGeographics pour les limites administratives
##      - DE: (C) EuroGeographics bezüglich der Verwaltungsgrenzen
##
##      For publications in languages other than
##      English, French or German,
##      the translation of the copyright notice
##      in the language of the publication shall be used.
##
##      If you intend to use the data commercially,
##      please contact EuroGeographics for
##      information regarding their licence agreements.
##
```

```
## data_frame at resolution 1: 10  cached at:  C:\Users\user\AppData\Local\Temp\Rtmpk3Zcbw/eurostat/df10.RData
```

```
# para dibujar solo paises stat_LEVEL_ 0:países 1:NUTS1 2:CCAA 3:Provincias
mapdatap<-mapdata%>%
  filter(STAT_LEVL_ == 0)

#ColorBr="YlOrRd","PuBuGn","RdYlBu","YlOrBr","Greens","RdPu","YlGn"
ggplot(mapdatap, aes(x = long, y = lat, group = group))+
  geom_polygon(aes(fill=cat), color="grey", size = .1)+
  scale_fill_brewer(palette = "YlOrBr") +
  labs(title="Peso En el PIB EUROPEO",
        subtitle="pps",
        fill="(%)") + theme_light()+
  coord_map(xlim=c(-12,35), ylim=c(35,67))  #35.67
```



#####

## 4 Tasas de Crecimiento del PIB

```
#Datos del PIB IV 2010=100
dat.pib<-dat.pib%>%
  filter(unit=="Chain linked volumes, index 2010=100")

dat.pib.a<-dat.pib%>%
  filter(time==2017)%>%
  mutate(PIB2017=values)%>%
  select(geo,geo_code,PIB2017)

dat.pib.b<-dat.pib%>%
  filter(time==2018)%>%
  mutate(PIB2018=values)%>%
  select(geo_code,PIB2018)

dat.pib.c<-inner_join(dat.pib.a,dat.pib.b,by="geo_code")

dat.pib.c<-dat.pib.c%>%
  mutate(Tpib=PIB2018/PIB2017-1)

levels(dat.pib.c$geo)
```

```
## [1] "Albania"
## [2] "Austria"
## [3] "Bosnia and Herzegovina"
## [4] "Belgium"
## [5] "Bulgaria"
## [6] "Switzerland"
## [7] "Cyprus"
## [8] "Czechia"
## [9] "Germany (until 1990 former territory of the FRG)"
## [10] "Denmark"
## [11] "Euro area (EA11-2000, EA12-2006, EA13-2007, EA15-2008, EA16-2010, EA17-2013, EA18-2014, EA19)"
## [12] "Euro area (12 countries)"
## [13] "Euro area (19 countries)"
## [14] "Estonia"
## [15] "Greece"
## [16] "Spain"
## [17] "European Union - 15 countries (1995-2004)"
## [18] "European Union - 27 countries (from 2019)"
## [19] "European Union - 28 countries"
## [20] "Finland"
## [21] "France"
## [22] "Croatia"
## [23] "Hungary"
## [24] "Ireland"
## [25] "Italy"
## [26] "Lithuania"
## [27] "Luxembourg"
## [28] "Latvia"
## [29] "North Macedonia"
## [30] "Netherlands"
## [31] "Norway"
## [32] "Poland"
## [33] "Portugal"
## [34] "Romania"
## [35] "Serbia"
## [36] "Sweden"
## [37] "Slovenia"
## [38] "Slovakia"
## [39] "United Kingdom"
## [40] "Iceland"
## [41] "Malta"
## [42] "Turkey"
## [43] "Montenegro"
## [44] "Kosovo (under United Nations Security Council Resolution 1244/99)"
## [45] "Liechtenstein"
```

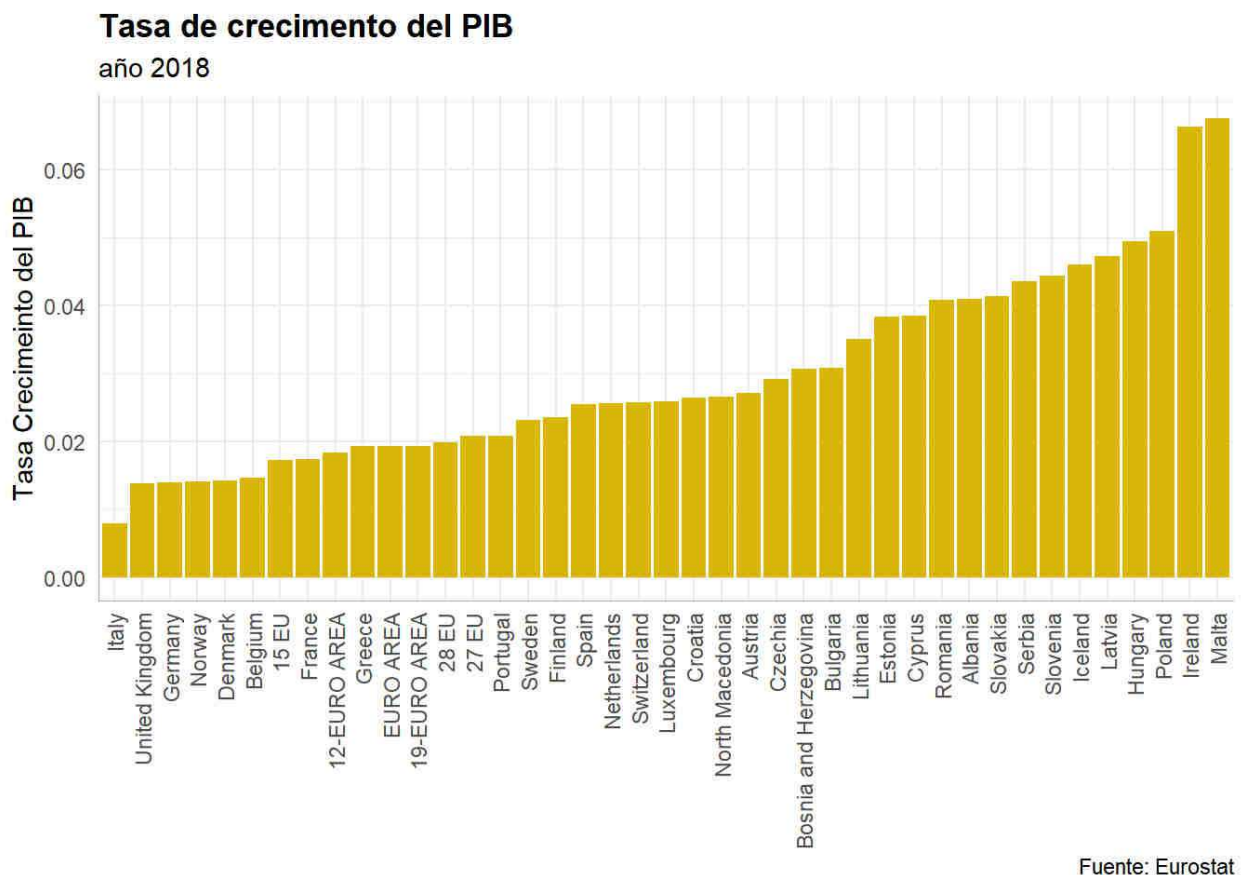
```

levels(dat.pib.c$geo)[9]<-"Germany"
levels(dat.pib.c$geo)[44]<-"Kosovo"
levels(dat.pib.c$geo)[17]<-"15 EU"
levels(dat.pib.c$geo)[18]<-"27 EU"
levels(dat.pib.c$geo)[19]<-"28 EU"
levels(dat.pib.c$geo)[11]<-"EURO AREA"
levels(dat.pib.c$geo)[12]<-"12-EURO AREA"
levels(dat.pib.c$geo)[13]<-"19-EURO AREA"

g2<-ggplot() +
  geom_bar(data=dat.pib.c, aes(x=reorder(geo, Tpib),y=Tpib) ,fill = "#D8B70A",stat = "identity")+
  ylab("Tasa Crecimeinto del PIB")+
  xlab(NULL)+
  labs( title="Tasa de crecimiento del PIB",
        subtitle="año 2018",
        caption = paste("Fuente: Eurostat", sep = "\n"))+
  theme_minimal()+
  theme(legend.position = "right",
        plot.title= element_text(colour="black", face="bold"),
        axis.line = element_line(size = 0.25, colour = "darkgrey", linetype = 1),
        legend.title=element_blank(),
        axis.text.x = element_text(angle = 90, hjust=1, vjust = 0.5))

print(g2)

```



```
#####
```

## 5 Tasas de Crecimiento de los Ocupados

```
#Datos de Ocupados
```

```
query<-search_eurostat("employment", type="table")
query
```

```
## # A tibble: 59 x 8
##   title code type `last update of~` `last table str~` `data start`
##   <chr> <chr> <chr> <chr> <chr> <chr>
## 1 Unem~ tgs0~ table 23.05.2019 23.05.2019 2007
## 2 Long~ tgs0~ table 23.05.2019 23.05.2019 2007
## 3 Youn~ tips~ table 25.04.2019 25.04.2019 2000
## 4 Tota~ tps0~ table 04.06.2019 04.06.2019 2007
## 5 Long~ sdg_~ table 24.04.2019 24.04.2019 2000
## 6 Harm~ teil~ table 25.06.2019 25.06.2019 2018M06
## 7 Harm~ teil~ table 25.06.2019 25.06.2019 2018M06
## 8 Harm~ teil~ table 25.06.2019 25.06.2019 2018M06
## 9 Harm~ teil~ table 25.06.2019 25.06.2019 2018M06
## 10 Harm~ teil~ table 25.06.2019 25.06.2019 2018M06
## # ... with 49 more rows, and 2 more variables: `data end` <chr>,
## # values <chr>
```

```
dat.ocu<-get_eurostat("nama_10_pe",time_format = "num",type = "code", stringsAsFactors = TRUE) #,time
_format = "raw",type = "code", stringsAsFactors = TRUE)
```

```
## Table nama_10_pe cached at C:\Users\user\AppData\Local\Temp\Rtmpk3Zcbw/eurostat/nama_10_pe_num_co
de_TF.rds
```

```
names(dat.ocu)
```

```
## [1] "unit" "na_item" "geo" "time" "values"
```

```
#Para ver que significan los campos
cc<-label_eurostat_vars(dat.ocu)
print(cc)
```

```
## [1] "Unit of measure"
## [2] "National accounts indicator (ESA 2010)"
## [3] "Geopolitical entity (reporting)"
## [4] "Period of time (a=annual, q=quarterly, m=monthly, d=daily, c=cumulated from January)"
```

```
# añadir los labels
dat.ocu<-label_eurostat(dat.ocu, code=names(dat.ocu))
```

```
table(dat.ocu$time)
```

```
##
## 1975 1976 1977 1978 1979 1980 1981 1982 1983 1984 1985 1986 1987 1988 1989
##   36   36   36   36   36   43   50   50   50   50   50   50   50   50   50
## 1990 1991 1992 1993 1994 1995 1996 1997 1998 1999 2000 2001 2002 2003 2004
##   53   63   70   77   84  277  470  470  471  472  482  492  492  492  492
## 2005 2006 2007 2008 2009 2010 2011 2012 2013 2014 2015 2016 2017 2018
##  492  493  494  499  504  504  504  504  515  526  526  526  500  472
```

```
dat.ocu<-dat.ocu%>%
  filter(time==2017 | time==2018)

table(dat.ocu$unit)
```

```
##
## Percentage change on previous period (based on persons)
##                                     486
##                               Thousand persons
##                                     486
```

```
dat.ocu<-dat.ocu%>%
  filter(unit=="Thousand persons")

table(dat.ocu$na_item)
```

```
##
## Total employment domestic concept Total employment national concept
##                               74                               62
## Total population national concept      Employees domestic concept
##                               80                               73
##      Employees national concept      Self-employed domestic concept
##                               62                               73
##      Self-employed national concept
##                               62
```

```
dat.ocu.a<-dat.ocu%>%
  filter(na_item_code=="EMP_DC")%>%
  filter(time==2017)%>%
  mutate(OCU2017=values)%>%
  select(geo,geo_code,OCU2017)

dat.ocu.b<-dat.ocu%>%
  filter(na_item_code=="EMP_DC")%>%
  filter(time==2018)%>%
  mutate(OCU2018=values)%>%
  select(geo_code,OCU2018)

dat.ocu.c<-inner_join(dat.ocu.a,dat.ocu.b,by="geo_code")

dat.ocu.c<-dat.ocu.c%>%
  mutate(Tocu=OCU2018/OCU2017-1)

levels(dat.ocu.c$geo)
```



```
## [1] "Austria"
## [2] "Belgium"
## [3] "Bulgaria"
## [4] "Switzerland"
## [5] "Cyprus"
## [6] "Czechia"
## [7] "Germany (until 1990 former territory of the FRG)"
## [8] "Denmark"
## [9] "Euro area (EA11-2000, EA12-2006, EA13-2007, EA15-2008, EA16-2010, EA17-2013, EA18-2014, EA19)"
## [10] "Euro area (12 countries)"
## [11] "Euro area (19 countries)"
## [12] "Estonia"
## [13] "Greece"
## [14] "Spain"
## [15] "European Union - 15 countries (1995-2004)"
## [16] "European Union - 27 countries (from 2019)"
## [17] "European Union - 28 countries"
## [18] "Finland"
## [19] "France"
## [20] "Hungary"
## [21] "Ireland"
## [22] "Iceland"
## [23] "Italy"
## [24] "Lithuania"
## [25] "Luxembourg"
## [26] "Latvia"
## [27] "Malta"
## [28] "Netherlands"
## [29] "Norway"
## [30] "Poland"
## [31] "Portugal"
## [32] "Romania"
## [33] "Sweden"
## [34] "Slovenia"
## [35] "Slovakia"
## [36] "United Kingdom"
## [37] "Serbia"
## [38] "Turkey"
## [39] "Croatia"
## [40] "Liechtenstein"
## [41] "Albania"
## [42] "Montenegro"
## [43] "North Macedonia"
```

```

levels(dat.ocu.c$geo)[7]<-"Germany"
levels(dat.ocu.c$geo)[15]<-"15 EU"
levels(dat.ocu.c$geo)[16]<-"27 EU"
levels(dat.ocu.c$geo)[17]<-"28 EU"
levels(dat.ocu.c$geo)[9]<-"EURO AREA"
levels(dat.ocu.c$geo)[10]<-"12-EURO AREA"
levels(dat.ocu.c$geo)[11]<-"19-EURO AREA"

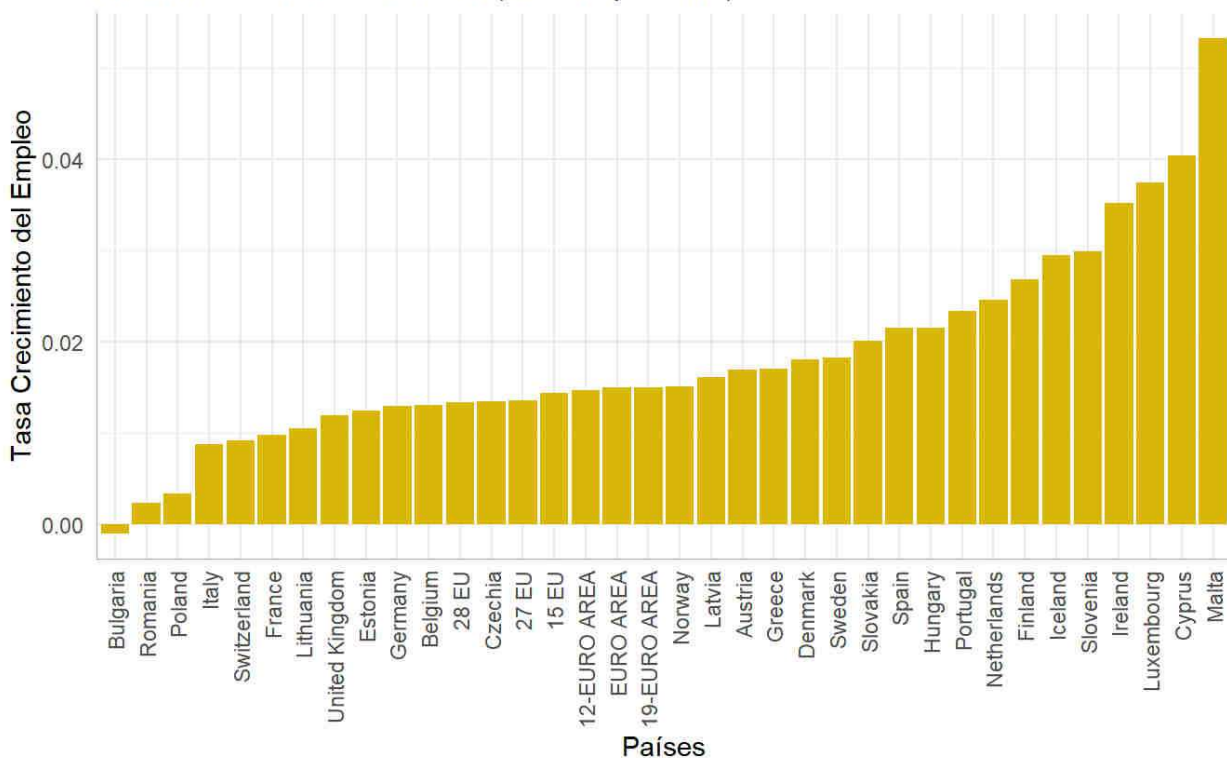
g3<-ggplot() +
  geom_bar(data=dat.ocu.c, aes(x=reorder(geo, Tocu),y=Tocu) ,fill = "#D8B70A",stat = "identity")+
  ylab("Tasa Crecimiento del Empleo")+
  xlab("Países")+
  labs( title="Tasa de crecimiento del Empleo",
        subtitle="año 2018. Contabilidad Nacional (miles de personas)",
        caption = paste("Fuente: Eurostat", sep = "\n"))+
  theme_minimal()+
  theme(legend.position = "right",
        plot.title= element_text(colour="black", face="bold"),
        axis.line = element_line(size = 0.25, colour = "darkgrey", linetype = 1),
        legend.title=element_blank(),
        axis.text.x = element_text(angle = 90, hjust=1, vjust = 0.5))

print(g3)

```

## Tasa de crecimiento del Empleo

año 2018. Contabilidad Nacional (miles de personas)



Fuente: Eurostat

## 6 Ley de Okun en la Unión Europea

```

# Fusiono las dos tablas

dat.EU<-inner_join(dat.ocu.c,dat.pib.c,by="geo_code")

```

```
## Warning: Column `geo_code` joining factors with different levels, coercing
## to character vector
```

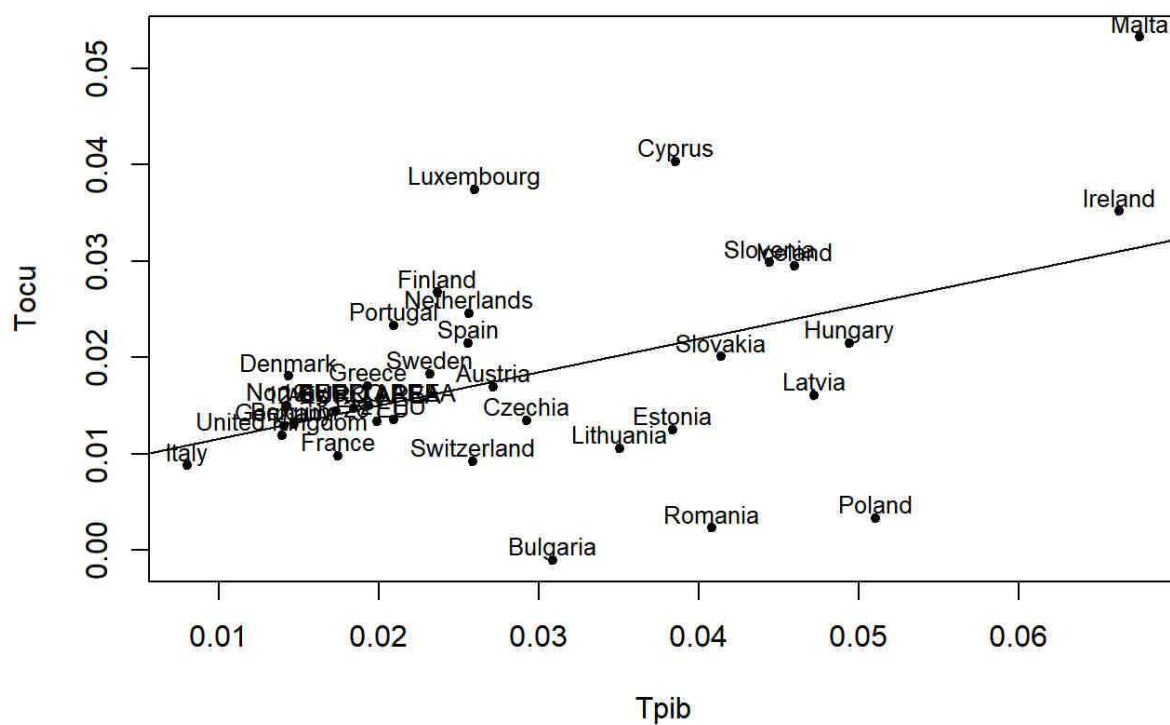
```
row.names(dat.EU)<-dat.EU$geo.x
```

```
## Warning: Setting row names on a tibble is deprecated.
```

```
dat.EU$pesoOCU<-dat.EU$OCU2018/as.numeric(dat.EU["28 EU","OCU2018"])
mediaEU<-mean(dat.EU$pesoOCU[-c(9,10,11,15,16,17)])
dat.EU$pesoOCU[c(9,10,11,15,16,17)]<-mediaEU
```

```
# Creo mapa ley de Okun
```

```
plot(Tocu~Tpib, data=dat.EU, pch=20 )
abline(lm(Tocu~Tpib, data=dat.EU))
text(x=dat.EU$Tpib, y=dat.EU$Tocu, labels = dat.EU$geo.x, cex=0.8, pos=3, offset = 0.15)
```



```

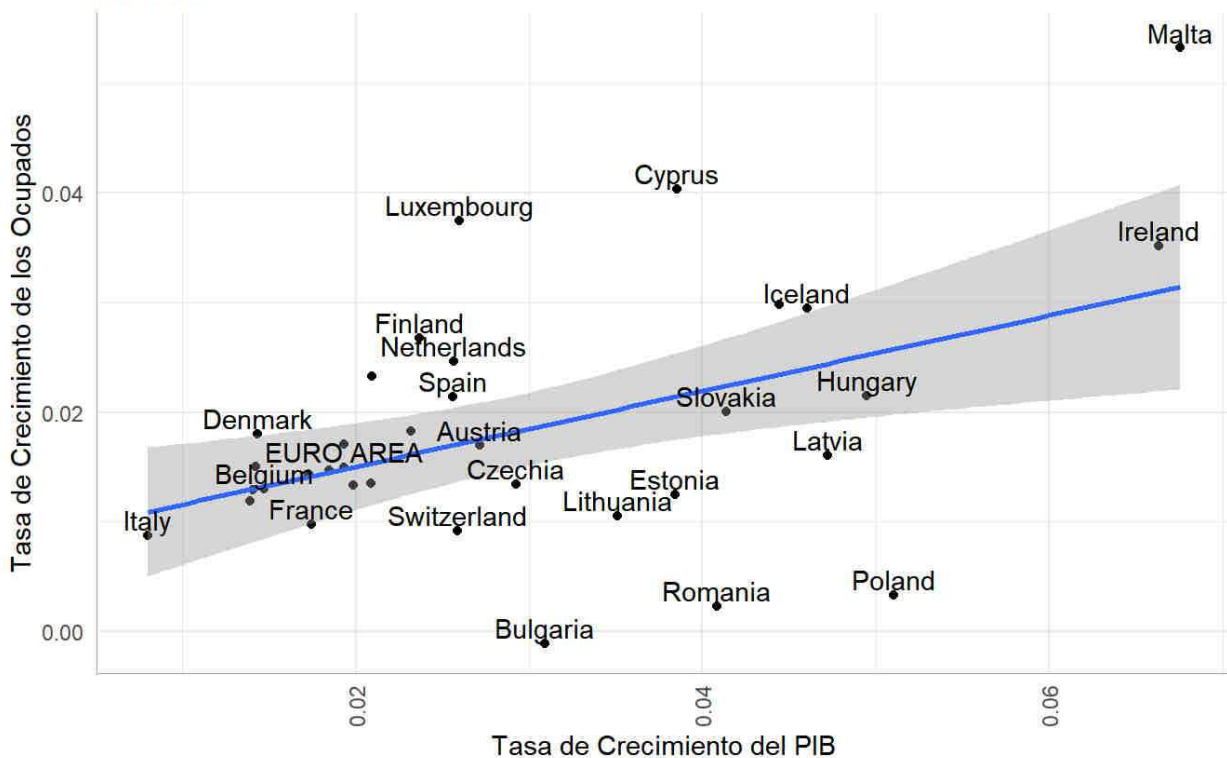
g4<-ggplot(data=dat.EU,aes(x=Tpib, y=Tocu, label=geo.x)) +
  geom_point()+
  geom_smooth(method='lm')+
  geom_text(check_overlap = TRUE, vjust = 0, nudge_y = 0.0005)+
  ylab("Tasa de Crecimiento de los Ocupados")+
  xlab("Tasa de Crecimiento del PIB")+
  labs( title="La Ley de Okun con países de Europa",
        subtitle=paste("Año 2018", sep = "\n"),
        caption = paste("Fuente: Elaboración Propia a partir de datos Eurostat", sep = "\n"))+
  theme_minimal()+
  theme(legend.position = "right",
        plot.title= element_text(colour="black", face="bold"),
        axis.line = element_line(size = 0.25, colour = "darkgrey", linetype = 1),
        legend.title=element_blank(),
        axis.text.x = element_text(angle = 90, hjust=1, vjust = 0.5))

print(g4)

```

## La Ley de Okun con países de Europa

Año 2018



Fuente: Elaboración Propia a partir de datos Eurostat

```

g4bis<-ggplot(data=dat.EU,aes(x=Tpib, y=Tocu, label=geo.x)) +
  geom_point(aes(size=pesoOCU))+
  geom_smooth(method='lm')+
  geom_text(aes(size=pesoOCU), check_overlap = TRUE, vjust = 0, nudge_y = 0.0005)+
  ylab("Tasa de Crecimiento de los Ocupados")+
  xlab("Tasa de Crecimiento del PIB")+
  labs( title="La Ley de Okun con países de Europa",
        subtitle=paste("Año 2018", "Tamaño en función de la importancia relativa en el total EU", se
p = "\n"),
        caption = paste("Fuente: Elaboración Propia a partir de datos Eurostat", sep = "\n"))+
  theme_minimal()+
  theme(legend.position = "right",
        plot.title= element_text(colour="black", face="bold"),
        axis.line = element_line(size = 0.25, colour = "darkgrey", linetype = 1),
        legend.title=element_blank(),
        axis.text.x = element_text(angle = 90, hjust=1, vjust = 0.5))

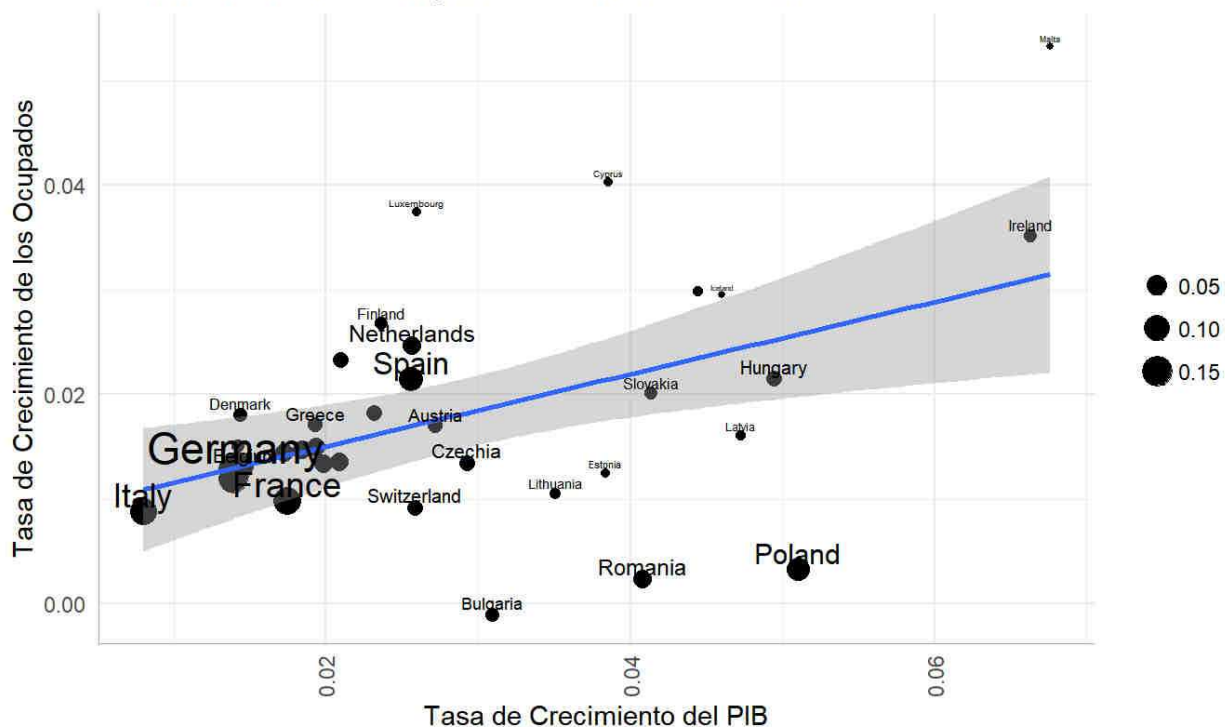
print(g4bis)

```

## La Ley de Okun con países de Europa

Año 2018

Tamaño en función de la importancia relativa en el total EU



Fuente: Elaboración Propia a partir de datos Eurostat

```
row.names(dat.EU)<-dat.EU$geo.x
```

```
## Warning: Setting row names on a tibble is deprecated.
```

```

okun.Europa<-lm(Tocu~Tpib, data=dat.EU)
stargazer(okun.Europa,type ="text",digits = 4,title = "Ley de Okun con datos de Europa" )

```

```
##
## Ley de Okun con datos de Europa
## =====
##                               Dependent variable:
##                               -----
##                               Tocu
## -----
## Tpib                        0.3449***
##                               (0.1125)
##
## Constant                    0.0081**
##                               (0.0037)
##
## -----
## Observations                36
## R2                          0.2167
## Adjusted R2                 0.1937
## Residual Std. Error        0.0098 (df = 34)
## F Statistic                 9.4069*** (df = 1; 34)
## =====
## Note:                       *p<0.1; **p<0.05; ***p<0.01
```

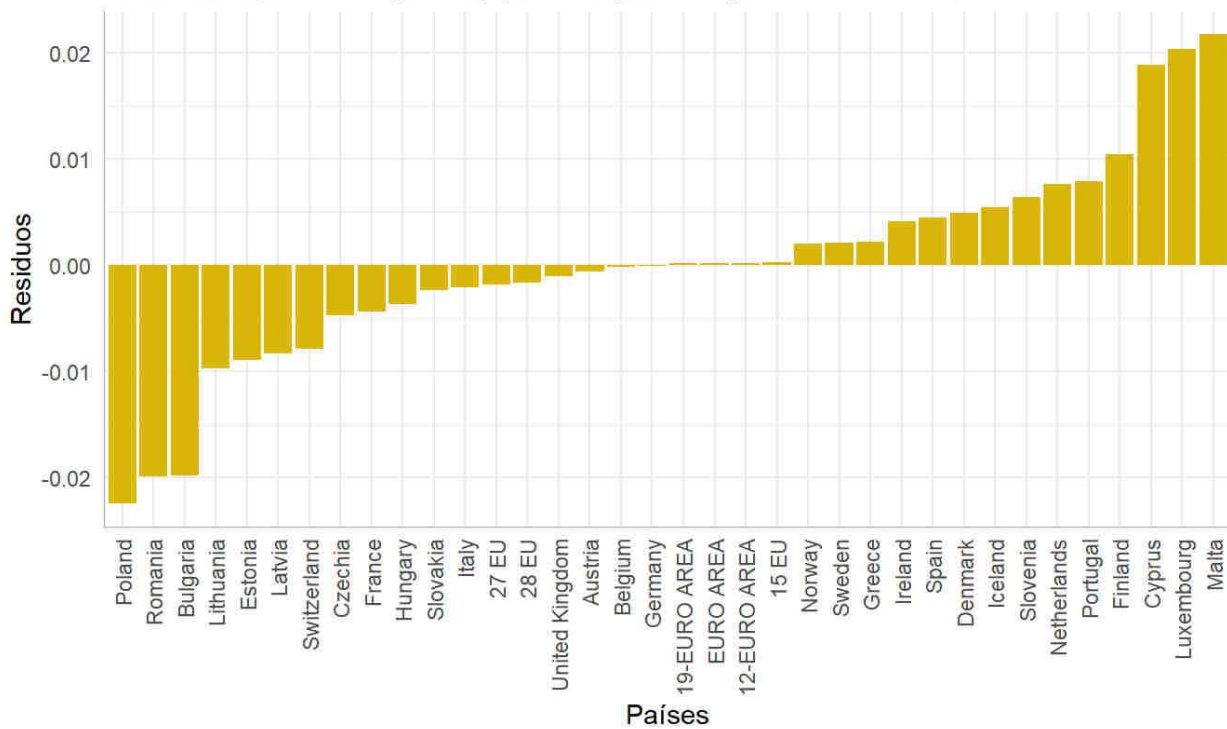
```
residuos.okun<-as.data.frame(residuals(okun.Europa))
names(residuos.okun)<-"residuo"

g5<-ggplot() +
  geom_bar(data=residuos.okun,aes(x=reorder(rownames(residuos.okun),residuo),y=residuo) ,fill = "#D8
B70A",stat = "identity")+
  ylab("Residuos")+
  xlab("Países")+
  labs( title="Residuos de la Ley de Okun con países de Europa",
        subtitle=paste("Valores Negativos implican crecimiento del empleo por debajo",
        "de la media (recta de regresión), y valores positivos por encima de la media", sep = "\n"
),
        caption = paste("Fuente: Eurostat", sep = "\n"))+
  theme_minimal()+
  theme(legend.position = "right",
        plot.title= element_text(colour="black", face="bold"),
        axis.line = element_line(size = 0.25, colour = "darkgrey", linetype = 1),
        legend.title=element_blank(),
        axis.text.x = element_text(angle = 90, hjust=1, vjust = 0.5))

print(g5)
```

## Residuos de la Ley de Okun con países de Europa

Valores Negativos implican crecimiento del empleo por debajo de la media (recta de regresión), y valores positivos por encima de la media



Fuente: Eurostat

```
#####3
```

```
row.names(dat.EU)<-dat.EU$geo_code
```

```
## Warning: Setting row names on a tibble is deprecated.
```

```
okun.Europa<-lm(Tocu~Tpib, data=dat.EU)
residuos.okun<-as.data.frame(residuals(okun.Europa))
names(residuos.okun)<-"residuo"
```

```
#Mapa
aamap<-residuos.okun%>%
  mutate(geo = rownames(residuos.okun))%>%
  mutate(cat = cut_to_classes(residuo, n=7,style="quantile", decimals=4))

mapdata <- merge_eurostat_geodata(aamap, all_regions = TRUE,
                                   resolution = "10")
```

```
##
##      COPYRIGHT NOTICE
##
##      When data downloaded from this page
##      <http://ec.europa.eu/eurostat/web/gisco/geodata/reference-data/administrative-units-statistical-units>
##      is used in any printed or electronic publication,
##      in addition to any other provisions
##      applicable to the whole Eurostat website,
##      data source will have to be acknowledged
##      in the legend of the map and
##      in the introductory page of the publication
##      with the following copyright notice:
##
##      - EN: (C) EuroGeographics for the administrative boundaries
##      - FR: (C) EuroGeographics pour les limites administratives
##      - DE: (C) EuroGeographics bezüglich der Verwaltungsgrenzen
##
##      For publications in languages other than
##      English, French or German,
##      the translation of the copyright notice
##      in the language of the publication shall be used.
##
##      If you intend to use the data commercially,
##      please contact EuroGeographics for
##      information regarding their licence agreements.
##
```

```
## Reading cache file C:\Users\user\AppData\Local\Temp\Rtmpk3Zcbw/eurostat/df10.RData
```

```
## data_frame at resolution 1: 10 read from cache file: C:\Users\user\AppData\Local\Temp\Rtmpk3Zcbw/eurostat/df10.RData
```

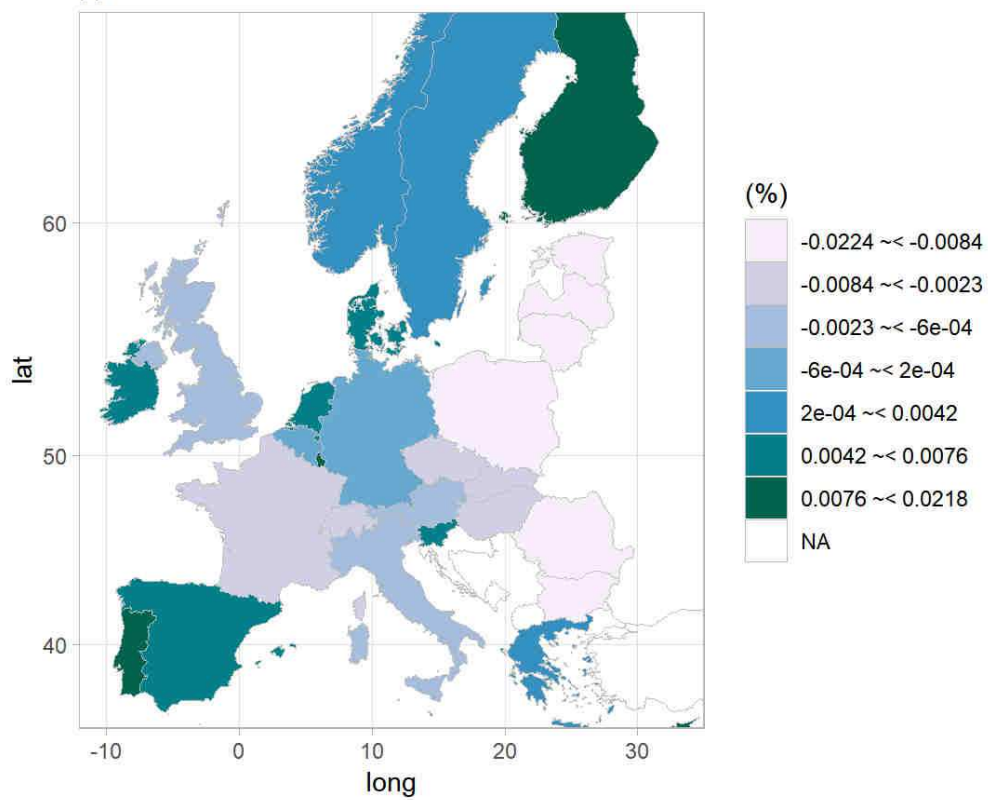
```
# para dibujar solo países stat_LEVEL_ 0:países 1:NUTS1 2:CCAA 3:Provincias
mapdatap<-mapdata%>%
  filter(STAT_LEVL_ == 0)

#ColorBr="YlOrRd","PuBuGn","RdYlBu","YlOrBr","Greens","RdPu","YlGn"
ggplot(mapdatap, aes(x = long, y = lat, group = group))+
  geom_polygon(aes(fill=cat), color="grey", size = .1)+
  scale_fill_brewer(palette = "PuBuGn") +
  labs(title="Residuos Ley Okun Europea",
        subtitle="pps",
        fill="(%)") + theme_light()+
  coord_map(xlim=c(-12,35), ylim=c(35,67)) #35.67
```



# Residuos Ley Okun Europea

pps



#####



Proyecto de innovación  
“Innova Docencia” de la UCM  
Convocatoria 2018-2019  
Proyecto número 275

**R-adaptación de la asignatura de métodos econométricos en economía y finanzas del  
Grado de Estadística Aplicada**

Facultad de Estudios Estadísticos  
Departamento de Economía Aplicada, Pública y Política

**Título**

***Práctica 5 : La ecuación de salarios y la discriminación por razón de género***

**Correspondiente al**

**Capítulo 2 REGRESIÓN LINEAL**

- El Modelo Lineal General (MLG)
- Hipótesis del modelo
- Estimación por Mínimos Cuadrados Ordinarios (MCO)
- Inferencia y predicción en el MLG
- Contrastes de cambio estructural
- Multicolinealidad y errores de medida

***Práctica 2 : La Ley de Okun: ¿a qué ritmo debe crecer la Economía para comenzar a generar empleo***

***Práctica 3 : La Ley de Okun: Test de Cambio Estructural, ¿han sido “eficaces” las reformas del mercado de trabajo?***

***Práctica 4 : Ley de Okun con datos de la UE desde la web de Eurostat***

***Práctica 5 : La ecuación de salarios y la discriminación por razón de género***

# Ecuación de Salarios y Discriminación Salarial de Género

*Asignatura*

*Métodos Econométricos en Economía y Finanzas*

*Profesores*

*Loren Escot, Alicia Pérez-Alonso y Julio E. Sandubete*

Dpto Economía Aplicada, Pública y Política

Facultad Estudios Estadísticos - Universidad Complutense de Madrid

*Curso académico 2018/2019*

## **Abstract**

En esta práctica se muestra que los coeficientes de un modelo lineal de una regresión en una ecuación de salarios pueden interpretarse como las remuneraciones que en media hace el mercado de cada una de las características de un individuo.

Se utilizará la descomposición de OAXACA-BLINDER para cuantificar la discriminación salarial contra la mujer, entendida como la diferente remuneración que reciben las mujeres de las mismas características que los hombres.

- 1 Tarea a entregar: ¿Están discriminadas las mujeres en nuestro mercado de trabajo?
- 2 Carga de datos sobre salarios
- 3 Análisis de la Brecha salarial
- 4 Ecuación salarial y remuneración de las características de los trabajadores
  - 4.1 Interpretación de coeficientes
  - 4.2 No linealidades
  - 4.3 Remuneración por cada hora adicional trabajada
  - 4.4 Remuneración por nivel de estudios
- 5 Modelo General de la ecuación de salarios
  - 5.1 Primeras evidencias de la discriminación de género
- 6 Cuantificación de la discriminación: Descomposición de Oaxaca-Blinder

## 1 Tarea a entregar: ¿Están discriminadas las mujeres en nuestro mercado de trabajo?

¿Están discriminadas las mujeres en nuestro mercado de trabajo? Tenéis que estimar al tamaño de la brecha salarial en España utilizando el fichero de salarios que utilizamos en la práctica de clase.

1. Utilizando una ecuación de salarios, tenéis que contrastar si a las mujeres se les remunera en media de manera diferente que a los varones
2. Utilizar la metodología de Oaxaca-Blinder para estimar que parte de la brecha salarial se debe a diferencias en características y que parte se debe a discriminación
3. Estimar cuanto tendría que aumentar el salario medio de las mujeres si fuesen remuneradas como los hombres
4. Estimar cuanto tendría que bajar el salario medio de los varones si éstos fuesen remunerados como las mujeres
5. Mostrar un histograma con la distribución de los indicadores de discriminación de las mujeres. # Librerías

```
library(excel.link)
```

```
## Warning: package 'excel.link' was built under R version 3.4.3
```

```
##  
## To Daniela Khazova who constantly inspires me...
```

```
library(dplyr)
```

```
## Warning: package 'dplyr' was built under R version 3.4.4
```

```
##  
## Attaching package: 'dplyr'
```

```
## The following objects are masked from 'package:stats':  
##  
## filter, lag
```

```
## The following objects are masked from 'package:base':  
##  
## intersect, setdiff, setequal, union
```

```
library(stargazer)
```

```
## Warning: package 'stargazer' was built under R version 3.4.1
```

```
##  
## Please cite as:
```

```
## Hlavac, Marek (2015). stargazer: Well-Formatted Regression and Summary Statistics Tables.
```

```
## R package version 5.2. http://CRAN.R-project.org/package=stargazer
```

## 2 Carga de datos sobre salarios

Datos sobre salarios de la encuesta de condiciones de vida del INE, 2010, para hacer prácticas. Microdatos sobre salarios extraídos de la Encuesta de Condiciones de Vida 2010 del INE:

[http://www.ine.es/dyngs/INEbase/es/operacion.htm?](http://www.ine.es/dyngs/INEbase/es/operacion.htm?c=Estadistica_C&cid=1254736176807&menu=resultados&secc=1254736195153&idp=1254735976608)

[c=Estadistica\\_C&cid=1254736176807&menu=resultados&secc=1254736195153&idp=1254735976608](http://www.ine.es/dyngs/INEbase/es/operacion.htm?c=Estadistica_C&cid=1254736176807&menu=resultados&secc=1254736195153&idp=1254735976608)

[http://www.ine.es/dyngs/INEbase/es/operacion.htm?](http://www.ine.es/dyngs/INEbase/es/operacion.htm?c=Estadistica_C&cid=1254736176807&menu=resultados&secc=1254736195153&idp=1254735976608)

[c=Estadistica\\_C&cid=1254736176807&menu=resultados&secc=1254736195153&idp=1254735976608](http://www.ine.es/dyngs/INEbase/es/operacion.htm?c=Estadistica_C&cid=1254736176807&menu=resultados&secc=1254736195153&idp=1254735976608))

```
# Lectura de datos un fichero excel  
Varones<-xl.read.file("salarios_Practica.xlsx", top.left.cell = "A1",header=TRUE, xl.sheet = "Hombres")  
Mujeres<-xl.read.file("salarios_Practica.xlsx", top.left.cell = "A1",header=TRUE, xl.sheet = "mujeres")
```

## 3 Análisis de la Brecha salarial

```
summary(Varones$SALARIO)
```

##	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
##	48	1150	1430	1645	1983	8667

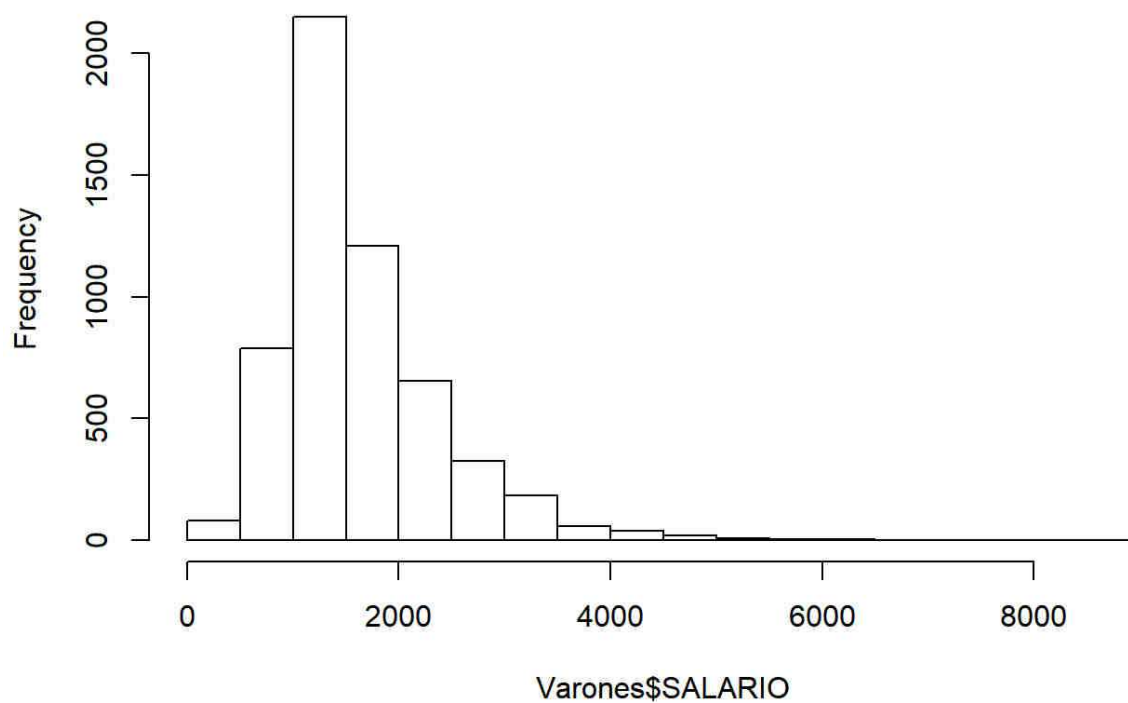
```
summary(Mujeres$SALARIO)
```

##	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
##	70.0	865.3	1166.7	1333.6	1680.0	6857.5

```
meanV<-mean(Varones$SALARIO)  
medianV<-median(Varones$SALARIO)
```

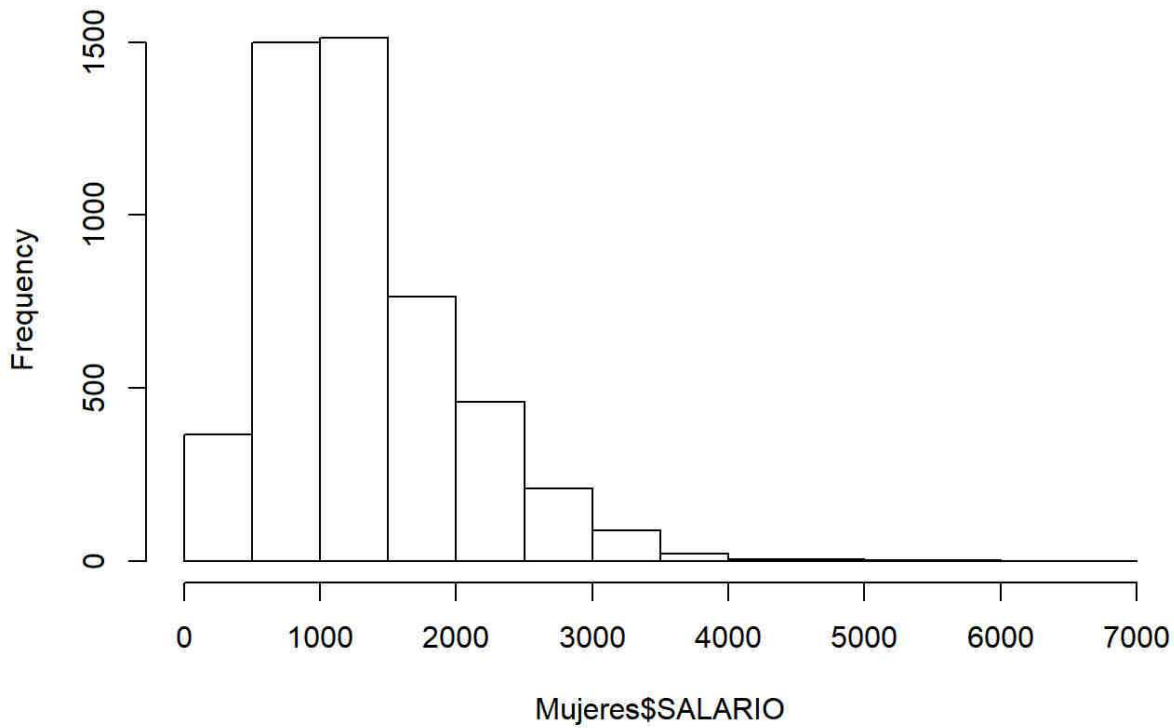
```
hist(Varones$SALARIO)
```

## Histogram of Varones\$SALARIO



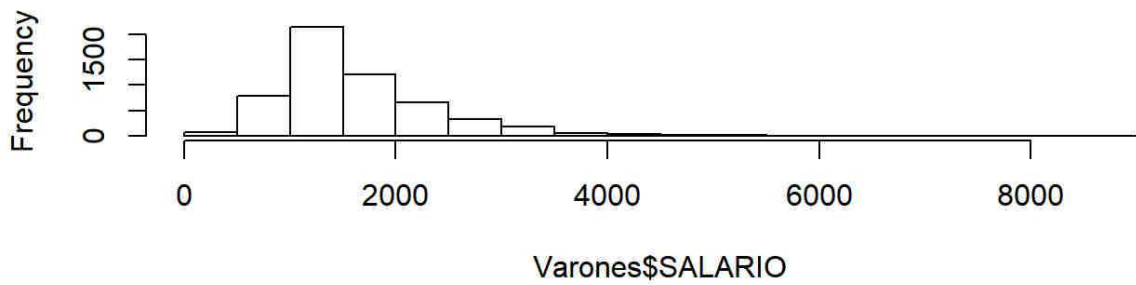
```
hist(Mujeres$SALARIO)
```

**Histogram of Mujeres\$SALARIO**

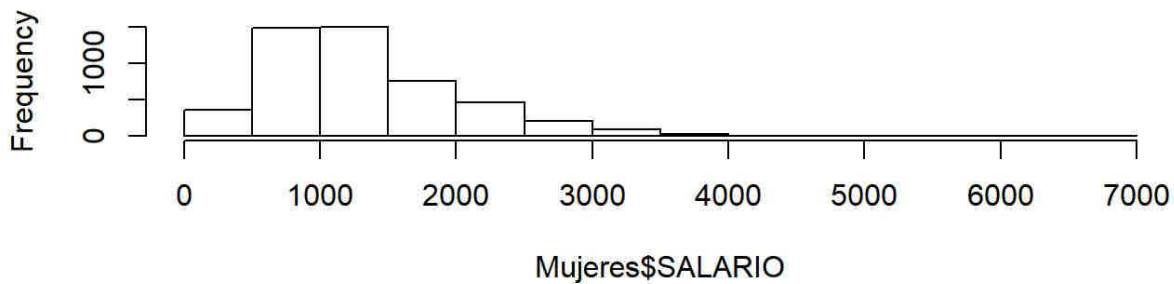


```
par(mfrow=c(2,1))  
hist(Varones$SALARIO)  
hist(Mujeres$SALARIO)
```

**Histogram of Varones\$SALARIO**



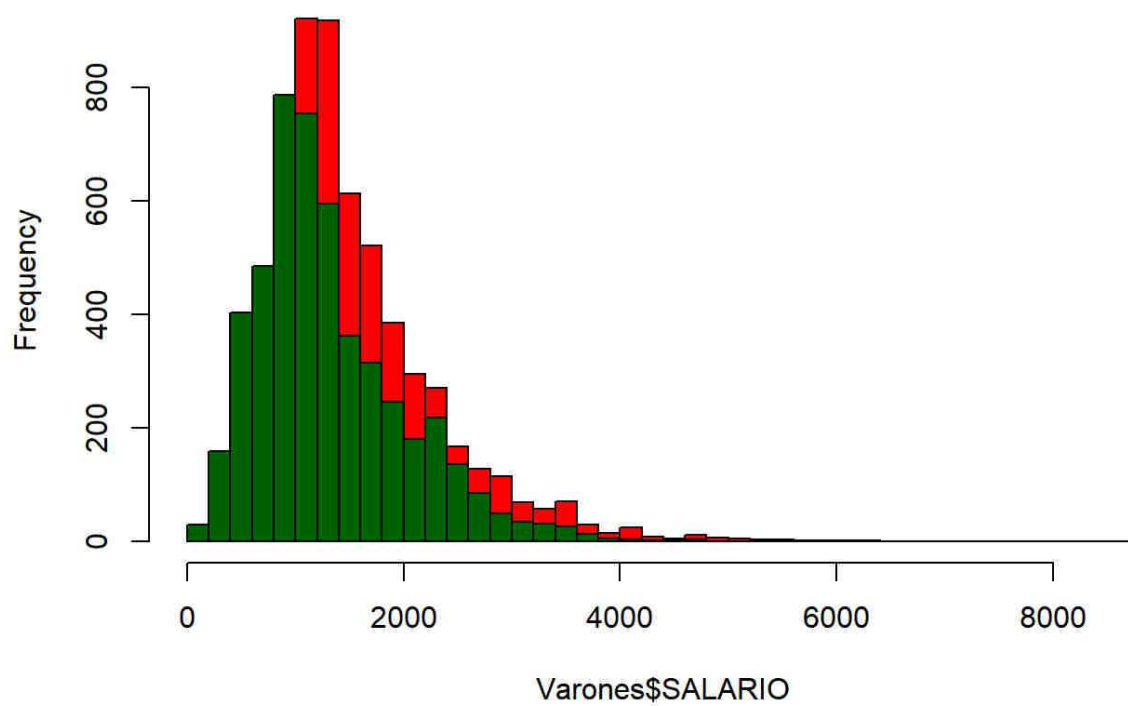
**Histogram of Mujeres\$SALARIO**



```
par(mfrow=c(1,1))

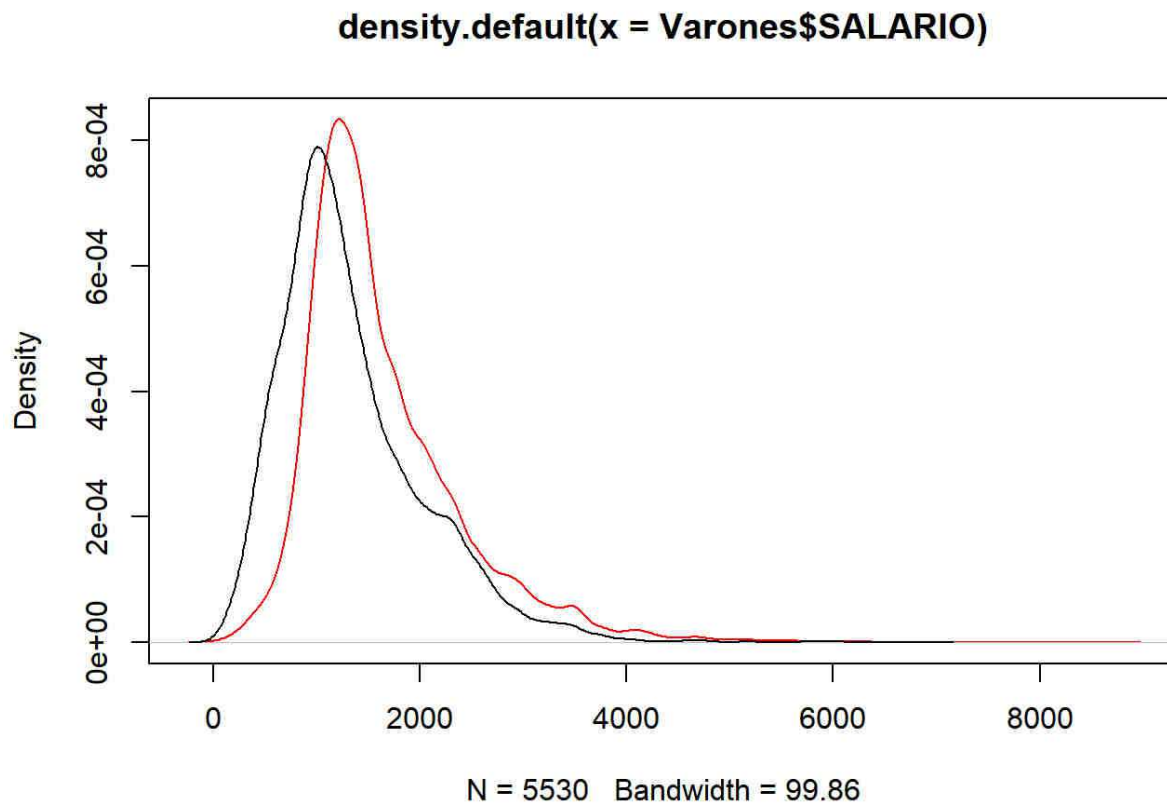
hist(Varones$SALARIO, col="red",nclass = 40)
hist(Mujeres$SALARIO, col="darkgreen",nclass = 40, add=TRUE)
```

### Histogram of Varones\$SALARIO



```
plot(density(Varones$SALARIO), col="red")
lines(density(Mujeres$SALARIO, col="darkgreen"))
```

```
## Warning: In density.default(Mujeres$SALARIO, col = "darkgreen") :
## extra argument 'col' will be disregarded
```



## 4 Ecuación salarial y remuneración de las características de los trabajadores

```
# Unimos los datos de Varones y mujeres
datos<-rbind.data.frame(Varones, Mujeres)
```

```
#Salarios medios por sexo
```

```
tapply(datos$SALARIO, datos$MUJER, mean) #summary
```

```
##          0          1
## 1645.131 1333.577
```

```
aa<-datos%>%
  group_by(MUJER)%>%
  summarise(mediaSalario=mean(SALARIO,na.rm = TRUE))
aa
```

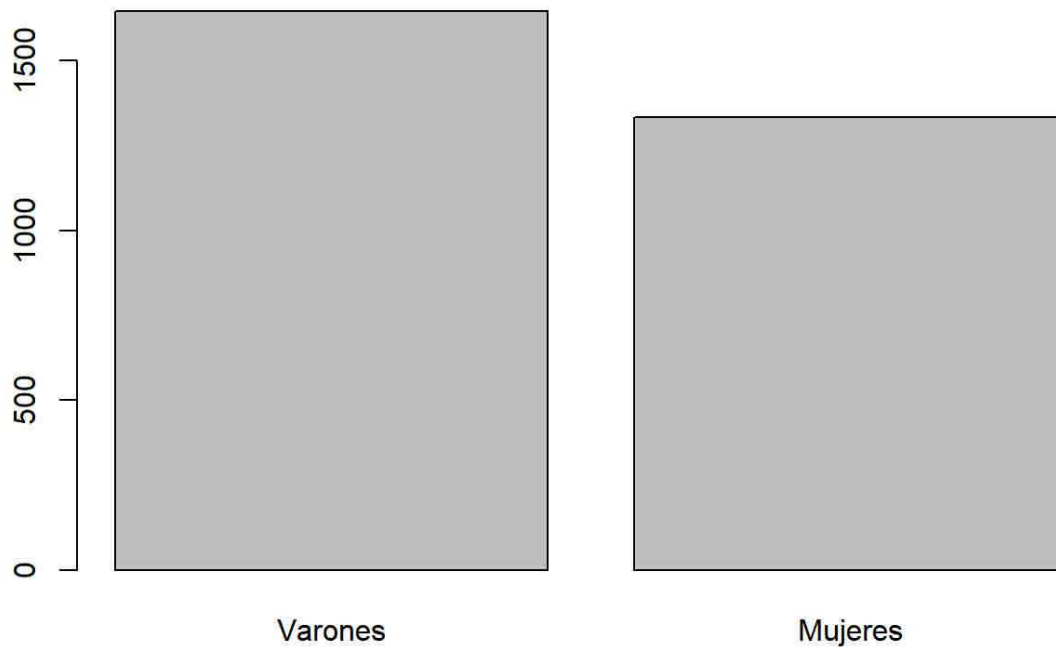
```
## # A tibble: 2 x 2
##   MUJER mediaSalario
##   <dbl>     <dbl>
## 1     0       1645.
## 2     1       1334.
```

```
# Brecha Salarial
aa[1,2]-aa[2,2]
```



```
## mediaSalario
## 1      311.5538
```

```
barplot(height = aa$mediaSalario, names.arg = c("Varones", "Mujeres"))
```



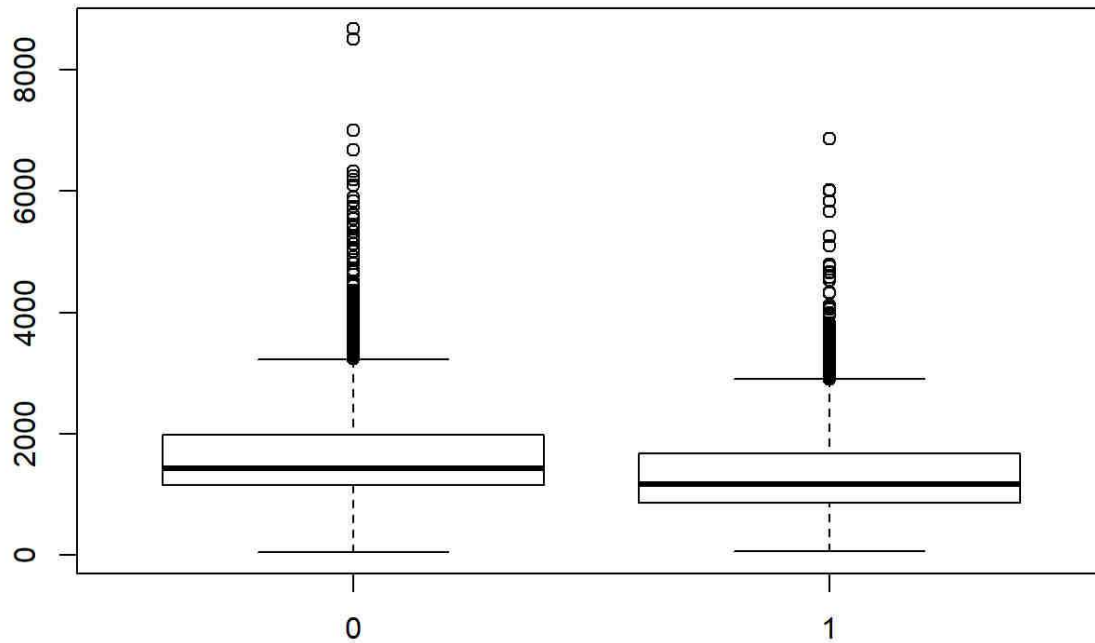
```
boxplot(SALARIO~MUJER, data=datos)
```

```
#test de diferencia de medias
```

```
oneway.test(SALARIO ~ MUJER, data = datos, var.equal = FALSE)
```

```
##
## One-way analysis of means (not assuming equal variances)
##
## data: SALARIO and MUJER
## F = 464.38, num df = 1, denom df = 10461, p-value < 2.2e-16
```

```
library(onewaytests) #para el test welch
```



```
library(RVAideMemoire) # para el test moods de diferencia de medianas
```

```
## *** Package RVAideMemoire v 0.9-65 ***
```

```
welch.out<-onewaytests::welch.test(SALARIO ~ MUJER, data=datos,na.rm=FALSE, verbose = F)
mood.out<-RVAideMemoire::mood.medtest(SALARIO ~ MUJER, data=datos,na.rm=FALSE, verbose = F)
```

```
welch.out$statistic
```

```
## [1] 464.3843
```

```
welch.out$p.value
```

```
## [1] 8.153282e-101
```

```
mood.out
```

```
##
## Mood's median test
##
## data: SALARIO by MUJER
## X-squared = 373.8, df = 1, p-value < 2.2e-16
```

## 4.1 Interpretación de coeficientes

```

modelo<-SALARIO ~ 1 + MUJER

modelo.lm<-lm(modelo, data=datos)

summary(modelo.lm)

```

```

##
## Call:
## lm(formula = modelo, data = datos)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1597.1  -483.6  -178.5   338.2  7021.5
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  1645.13      9.98   164.84  <2e-16 ***
## MUJER        -311.55     14.53   -21.44  <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 742.2 on 10467 degrees of freedom
## Multiple R-squared:  0.04207,    Adjusted R-squared:  0.04198
## F-statistic: 459.7 on 1 and 10467 DF,  p-value: < 2.2e-16

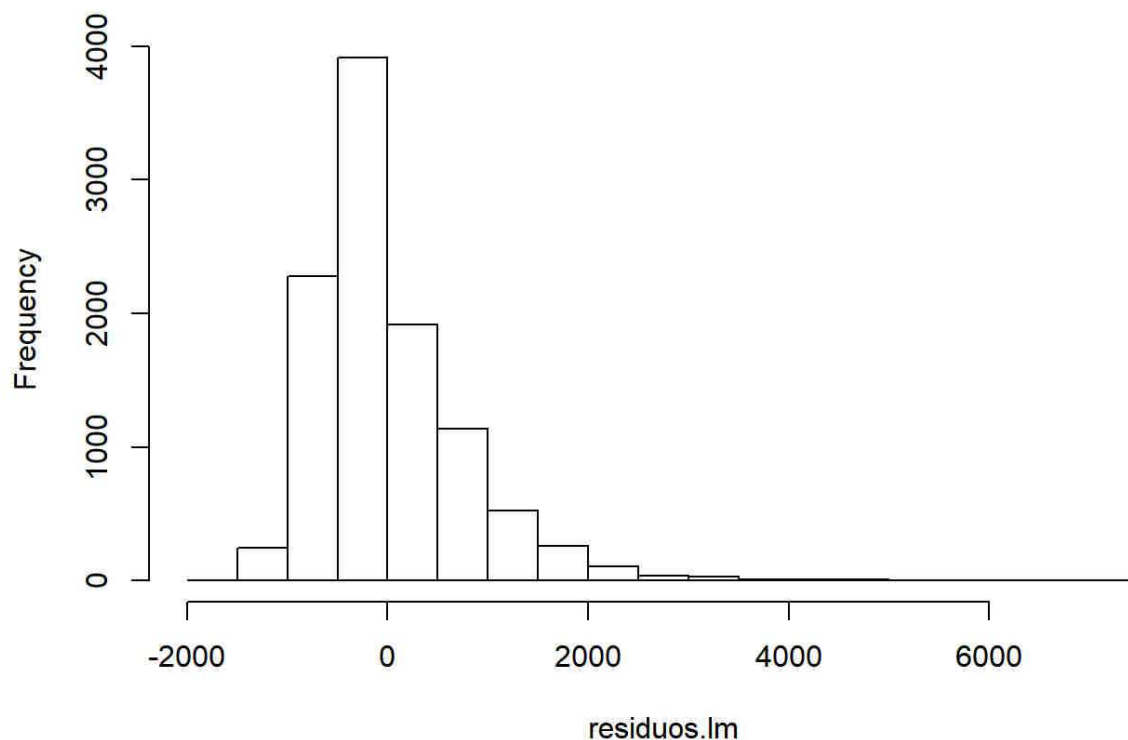
```

```

residuos.lm<-residuals(modelo.lm)
hist(residuos.lm)

```

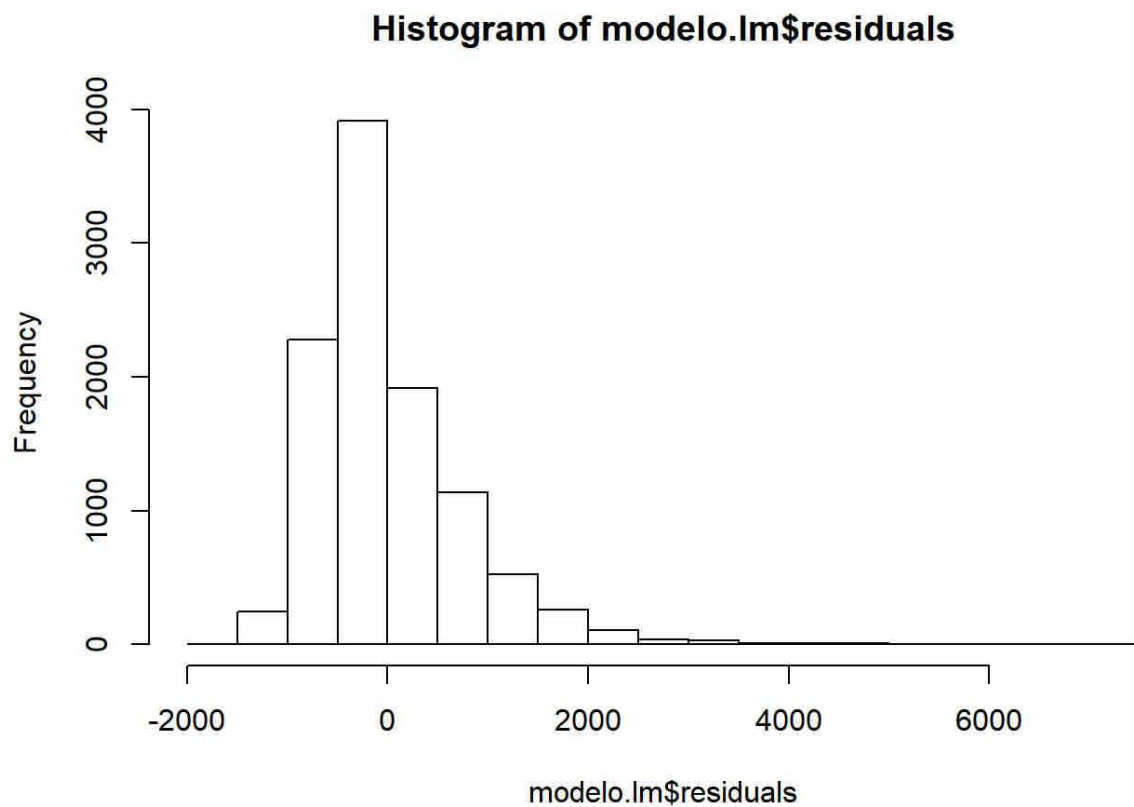
**Histogram of residuos.lm**



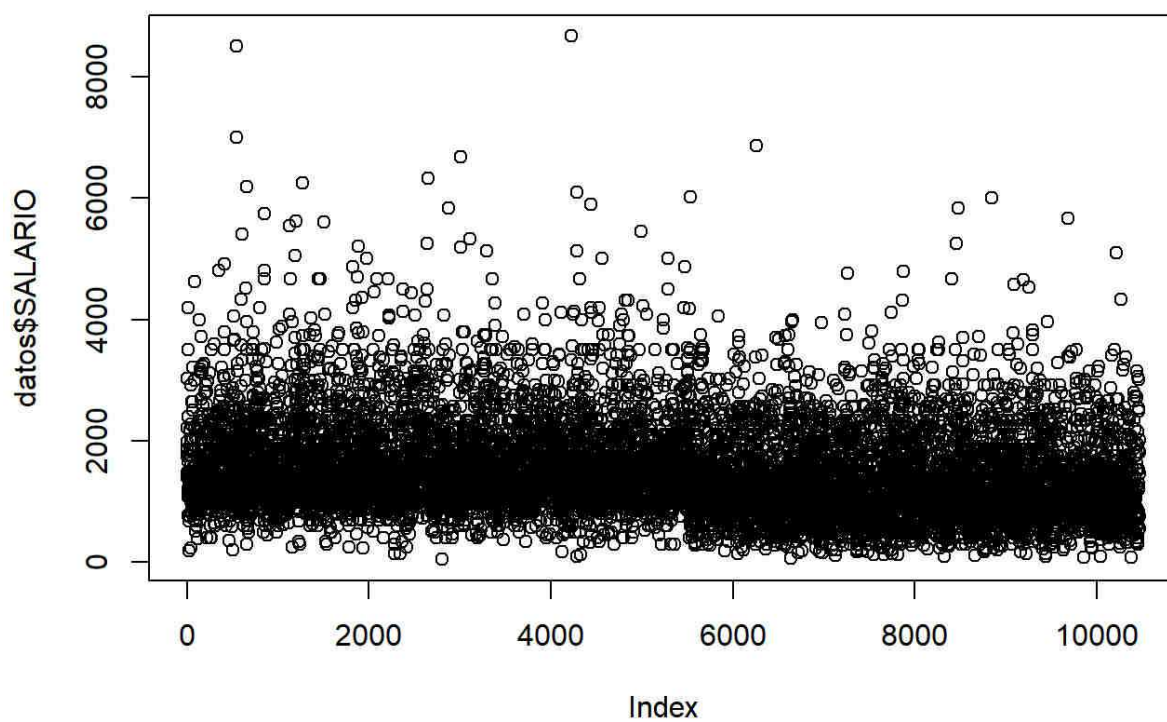
```

hist(modelo.lm$residuals)

```



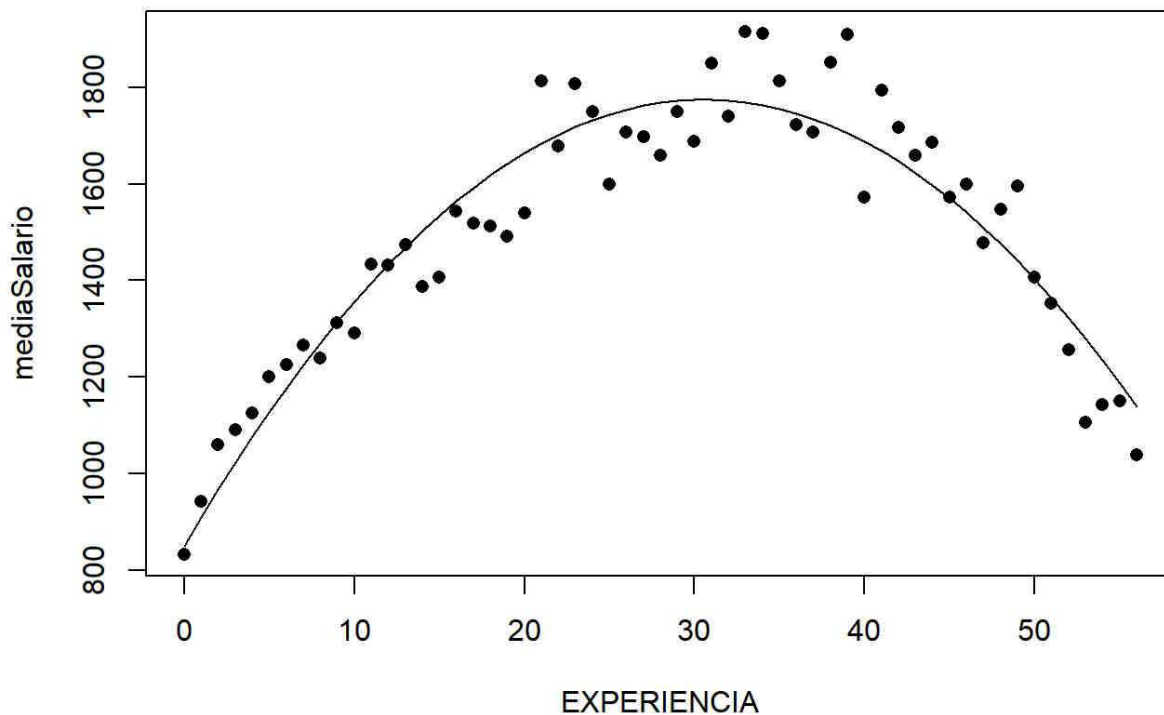
```
plot(datos$SALARIO)
```



## 4.2 No linealidades

```
aa<-datos%>%
  group_by(EXPERIENCIA)%>%
  summarise(mediaSalario=mean(SALARIO,na.rm = TRUE))

plot(mediaSalario ~ EXPERIENCIA, pch=16, data=aa)
lines(fitted(lm(mediaSalario ~ EXPERIENCIA + I(EXPERIENCIA^2), data=aa)) ~ EXPERIENCIA, data=aa)
```



```
coef(lm(mediaSalario ~ EXPERIENCIA + I(EXPERIENCIA^2), data=aa))
```

##	(Intercept)	EXPERIENCIA	I(EXPERIENCIA^2)
##	850.3771874	60.4229853	-0.9873204

## 4.3 Remuneración por cada hora adicional trabajada

```
# Modelo lm

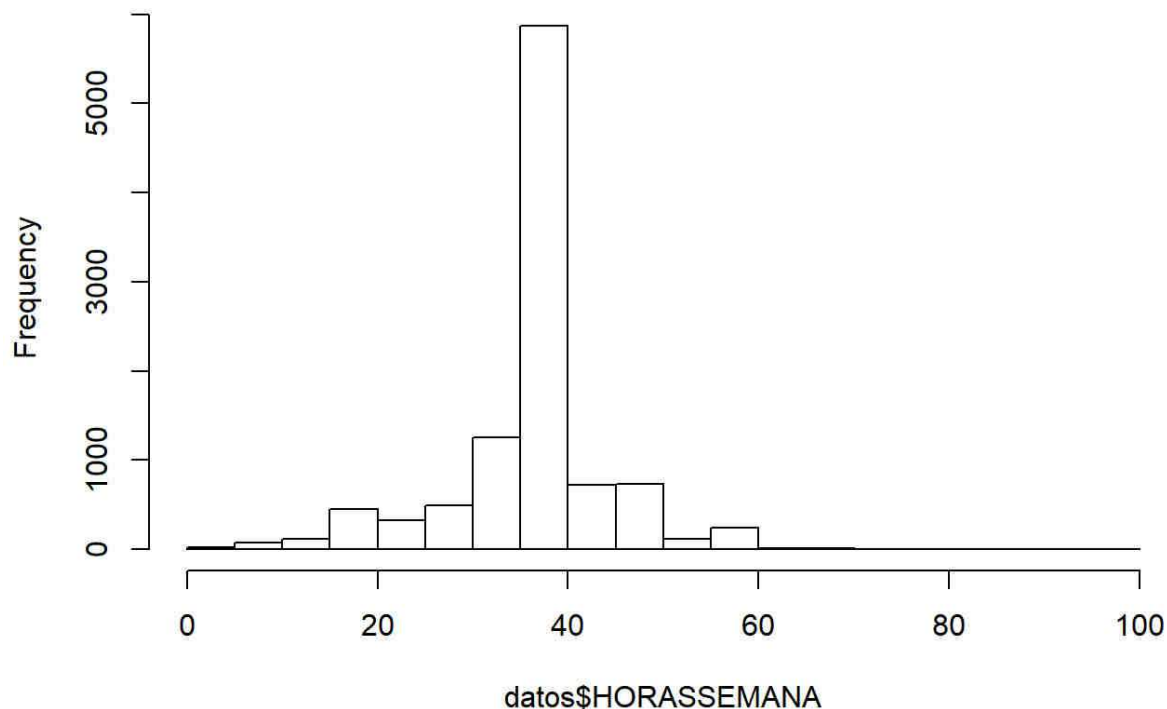
modelo<-SALARIO ~ HORASSEMANA
model1.lm<-lm(modelo, data=datos)
summary(model1.lm)
```

```
##
## Call:
## lm(formula = modelo, data = datos)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2305.3  -489.9  -200.7   329.1  6571.9
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  437.4822    32.4984   13.46  <2e-16 ***
## HORASSEMANA  27.6213     0.8262   33.43  <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 720.8 on 10467 degrees of freedom
## Multiple R-squared:  0.09648,    Adjusted R-squared:  0.0964
## F-statistic: 1118 on 1 and 10467 DF,  p-value: < 2.2e-16
```

```
# HORAS
```

```
hist(datos$HORASSEMANA)
```

### Histogram of datos\$HORASSEMANA



```
summary(datos$HORASMENSUALES)
```

```
## Length Class Mode
##      0  NULL  NULL
```

```
datos$HORASMENSUALES<-datos$HORASSEMANA*52/12
```

```
# Modelo lm
```

```
modelo<-SALARIO ~ HORASMENSUALES
model2.lm<-lm(modelo, data=datos)
summary(model2.lm)
```

```
##
## Call:
## lm(formula = modelo, data = datos)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2305.3  -489.9  -200.7   329.1  6571.9
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)   437.4822    32.4984   13.46  <2e-16 ***
## HORASMENSUALES  6.3742     0.1907   33.43  <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 720.8 on 10467 degrees of freedom
## Multiple R-squared:  0.09648,    Adjusted R-squared:  0.0964
## F-statistic: 1118 on 1 and 10467 DF,  p-value: < 2.2e-16
```

```
stargazer(modelo.lm,model1.lm,model2.lm, type="text")
```

```
##
## =====
##                               Dependent variable:
##                               -----
##                               SALARIO
##                               (1)      (2)      (3)
## -----
## MUJER                        -311.554***
##                               (14.530)
##
## HORASSEMANA                   27.621***
##                               (0.826)
##
## HORASMENSUALES                6.374***
##                               (0.191)
##
## Constant                     1,645.131***  437.482***  437.482***
##                               (9.980)    (32.498)    (32.498)
##
## -----
## Observations                 10,469      10,469      10,469
## R2                           0.042       0.096       0.096
## Adjusted R2                  0.042       0.096       0.096
## Residual Std. Error (df = 10467) 742.179    720.793    720.793
## F Statistic (df = 1; 10467)    459.735*** 1,117.725*** 1,117.725***
## =====
## Note:                         *p<0.1; **p<0.05; ***p<0.01
```

## 4.4 Remuneración por nivel de estudios

```
#ESTUDIOS
tabla_ESup<-datos%>%
  group_by(ESUPERIORES)%>%
  summarise(mean(SALARIO))

print(tabla_ESup)
```

```
## # A tibble: 2 x 2
##   ESUPERIORES `mean(SALARIO)`
##   <dbl>         <dbl>
## 1      0      1241.
## 2      1      1872.
```

```
#datos$ESUPERIORES<-as.factor(datos$ESUPERIORES)
#datos$ESECUNDARIOS<-as.factor(datos$ESECUNDARIOS)
#datos$EPRIMARIOS<-as.factor(datos$EPRIMARIOS)

modelo<-SALARIO ~ ESUPERIORES + ESECUNDARIOS + EPRIMARIOS

model3.lm<-lm(modelo, data=datos)
#summary(model3.lm)

modelo<-SALARIO ~ ESUPERIORES + ESECUNDARIOS + EPRIMARIOS - 1
model3b.lm<-lm(modelo, data=datos)

stargazer(model3.lm, model3b.lm, type="text")
```

##	=====		
##	Dependent variable:		
##	-----		
##	SALARIO		
##	(1)	(2)	
##	-----		
##	ESUPERIORES	749.262***	1,871.621***
##		(22.483)	(10.575)
##			
##	ESECUNDARIOS	147.821***	1,270.181***
##		(22.118)	(9.777)
##			
##	EPRIMARIOS		1,122.359***
##			(19.840)
##			
##	Constant	1,122.359***	
##		(19.840)	
##			
##	-----		
##	Observations	10,469	10,469
##	R2	0.170	0.831
##	Adjusted R2	0.170	0.831
##	Residual Std. Error (df = 10466)	690.715	690.715
##	F Statistic	1,074.817*** (df = 2; 10466)	17,133.580*** (df = 3; 10466)
##	=====		
##	Note:	*p<0.1; **p<0.05; ***p<0.01	



```

modelo<-SALARIO ~ MUJER+ ESUPERIORES + ESECUNDARIOS + EPRIMARIOS
model3c.lm<-lm(modelo, data=datos)

modelo<-SALARIO ~ MUJER+ ESUPERIORES + ESECUNDARIOS + EPRIMARIOS -1
model3d.lm<-lm(modelo, data=datos)

modelo<-SALARIO ~ MUJER+ ESUPERIORES + ESECUNDARIOS + EPRIMARIOS + MUJER:ESUPERIORES + MUJER:ESECUNDARIOS + MUJER:EPRIMARIOS
model3e.lm<-lm(modelo, data=datos)

stargazer(model3c.lm,model3e.lm, type="text")

```

```

##
## =====
##                               Dependent variable:
##                               -----
##                               SALARIO
##                               (1)                (2)
## -----
## MUJER                -396.160***             -420.837***
##                      (13.071)                (39.315)
##
## ESUPERIORES          816.916***             787.552***
##                      (21.673)                (28.344)
##
## ESECUNDARIOS          171.440***             173.579***
##                      (21.223)                (27.069)
##
## EPRIMARIOS
##
## MUJER:ESUPERIORES                    61.665
##                                     (44.273)
##
## MUJER:ESECUNDARIOS                    -1.541
##                                     (43.628)
##
## MUJER:EPRIMARIOS
##
## Constant              1,270.429***           1,279.652***
##                      (19.641)                (24.036)
## -----
## Observations              10,469              10,469
## R2                        0.237                0.238
## Adjusted R2              0.237                0.237
## Residual Std. Error    662.293 (df = 10465)    662.178 (df = 10463)
## F Statistic            1,085.563*** (df = 3; 10465) 652.686*** (df = 5; 10463)
## =====
## Note:                    *p<0.1; **p<0.05; ***p<0.01

```

## 5 Modelo General de la ecuación de salarios

```
# MODELO GENERAL
# con mujeres
modelo<-SALARIO ~ MUJER + ESUPERIORES + ESECUNDARIOS + EPRIMARIOS +
  +EXPERIENCIA + I(EXPERIENCIA^2) + ENFERMEDAD+ HORASMENSUALES+ SUPERVISOR+
  TEMPORAL+ EXTRANJERO + CASADO+ CASADO*MUJER+ Numerohijos12

model4.lm<-lm(modelo, data=datos)
#summary(model4.lm)

stargazer(model4.lm, type="text")
```

##	=====
##	Dependent variable:
##	-----
##	SALARIO
##	-----
##	MUJER
##	-113.815***
##	(19.266)
##	
##	ESUPERIORES
##	822.984***
##	(19.699)
##	
##	ESECUNDARIOS
##	226.236***
##	(18.732)
##	
##	EPRIMARIOS
##	
##	
##	EXPERIENCIA
##	26.425***
##	(1.880)
##	
##	I(EXPERIENCIA2)
##	-0.254***
##	(0.041)
##	
##	ENFERMEDAD
##	-39.692***
##	(14.404)
##	
##	HORASMENSUALES
##	4.337***
##	(0.160)
##	
##	SUPERVISOR
##	321.834***
##	(13.825)
##	
##	TEMPORAL
##	-132.529***
##	(14.631)
##	
##	EXTRANJERO
##	-180.181***
##	(19.962)
##	
##	CASADO
##	150.817***
##	(19.008)
##	
##	Numerohijos12
##	20.628**
##	(8.298)
##	
##	MUJER:CASADO
##	-105.233***
##	(24.063)
##	
##	Constant
##	-95.574**
##	(37.891)
##	
##	-----
##	Observations
##	10,469
##	R2
##	0.449
##	Adjusted R2
##	0.449
##	Residual Std. Error
##	563.096 (df = 10455)
##	F Statistic
##	655.926*** (df = 13; 10455)
##	=====
##	Note:
##	*p<0.1; **p<0.05; ***p<0.01

## 5.1 Primeras evidencias de la discriminación de genero

```
# Evidencia de discriminación, la remuneración (coeficientes)
# que se hace a varones y mujeres es diferente

# MODELO GENERAL
# SIN mujeres
modelo<-SALARIO ~ ESUPERIORES + ESECUNDARIOS + EPRIMARIOS +
  +EXPERIENCIA + I(EXPERIENCIA^2) + ENFERMEDAD+ HORASMENSUALES+ SUPERVISOR+
  TEMPORAL+ EXTRANJERO + CASADO+ Numerohijos12

model.lm<-lm(modelo, data=datos)
#summary(model.lm)
stargazer(model.lm, type="text")
```

```

##
## =====
##                               Dependent variable:
##                               -----
##                               SALARIO
## -----
## ESUPERIORES                801.030***
##                             (19.875)
##
## ESECUNDARIOS                223.291***
##                             (18.952)
##
## EPRIMARIOS
##
## EXPERIENCIA                 26.541***
##                             (1.901)
##
## I(EXPERIENCIA2)             -0.220***
##                             (0.042)
##
## ENFERMEDAD                  -46.139***
##                             (14.564)
##
## HORASMENSUALES              5.078***
##                             (0.155)
##
## SUPERVISOR                  342.748***
##                             (13.921)
##
## TEMPORAL                    -123.614***
##                             (14.792)
##
## EXTRANJERO                  -185.375***
##                             (20.193)
##
## CASADO                      94.289***
##                             (13.956)
##
## Numerohijos12               28.696***
##                             (8.354)
##
## Constant                    -283.913***
##                             (35.642)
##
## -----
## Observations                10,469
## R2                           0.436
## Adjusted R2                  0.435
## Residual Std. Error        569.748 (df = 10457)
## F Statistic                 734.943*** (df = 11; 10457)
## =====
## Note:                        *p<0.1; **p<0.05; ***p<0.01

```

```

## H0: AUSENCIA DE CAMBIO ESTRUCTURAL
library(strucchange)

```

```

## Warning: package 'strucchange' was built under R version 3.4.3

```

```
## Loading required package: zoo
```

```
##  
## Attaching package: 'zoo'
```

```
## The following objects are masked from 'package:base':  
##  
## as.Date, as.Date.numeric
```

```
## Loading required package: sandwich
```

```
# Test de Cambio structural Chow  
  
Chowtest<-Fstats(modelo, from=5531, to=5531, data=datos)  
Chowtest$Fstats
```

```
## Time Series:  
## Start = c(0, 5532)  
## End = c(0, 5532)  
## Frequency = 10469  
## [1] 346.3131
```

```
sctest(Chowtest, type="expF")
```

```
##  
## expF test  
##  
## data: Chowtest  
## exp.F = 173.16, p-value < 2.2e-16
```

```
sctest(Chowtest, type="aveF")
```

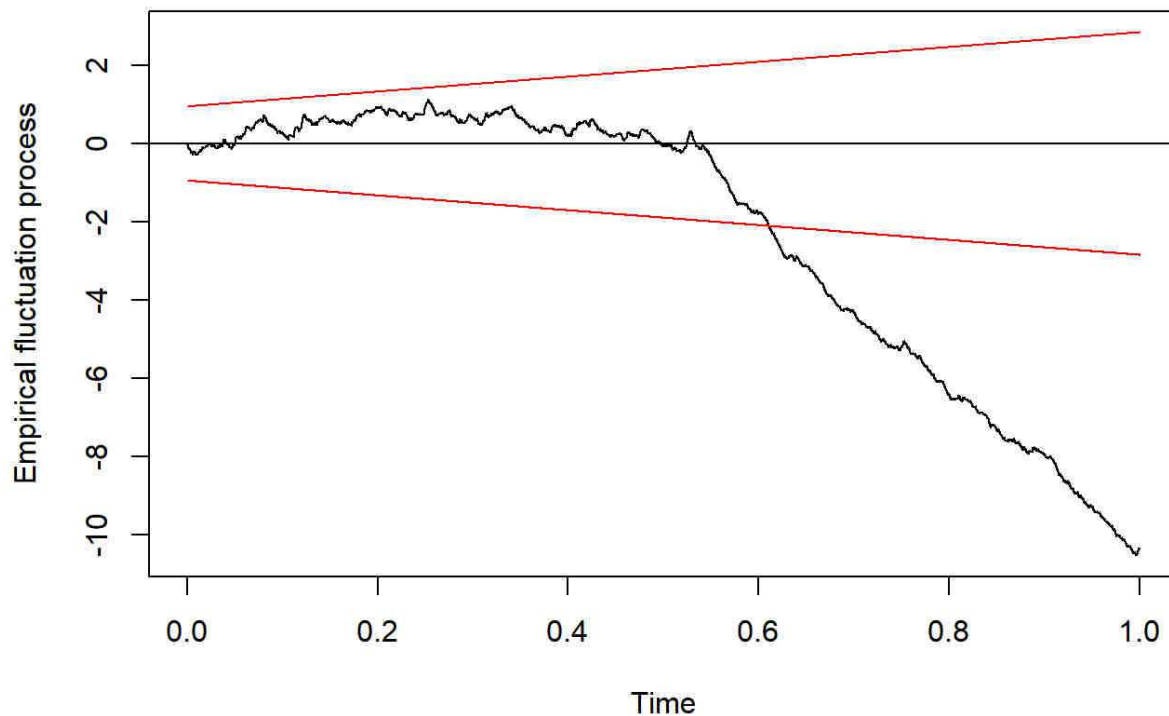
```
##  
## Chow test  
##  
## data: Chowtest  
## ave.F = 346.31, p-value < 2.2e-16
```

```
sctest(Chowtest, type="supF")
```

```
##  
## Chow test  
##  
## data: Chowtest  
## sup.F = 346.31, p-value < 2.2e-16
```

```
testestbr<-efp(modelo, type="Rec-CUSUM", data = datos, h=0.2)  
plot(testestbr)
```

## Recursive CUSUM test



```
#puntos donde parece que hay un cambio estructural
#locator(10, type = "p" ,col=2, pch=19)
0.5274800*nrow(datos) # número aprox de cambio (Coincide con el inici de las mujeres)
```

```
## [1] 5522.188
```

```
0.6295250*nrow(datos) # aquí es donde sale de bandas
```

```
## [1] 6590.497
```

## 6 Cuantificación de la discriminación: Descomposición de Oaxaca-Blinder

```
#primera evidencia de que hay cambio estructural
#entre los varones no hay cambio estructural
#habria que aplicar este test de cambio estructural por una parte a las mujeres y por otra a los hom
bres

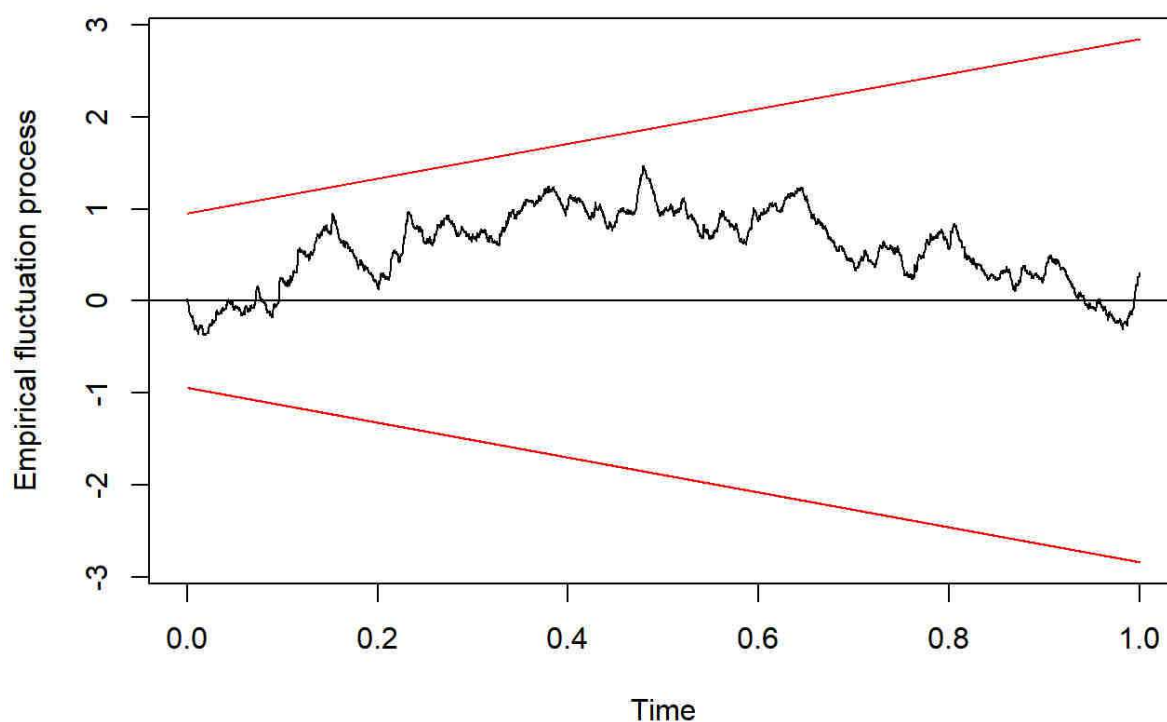
library(stargazer)
#MODELO VARONES:
Varones$HORASMENSALES <- Varones$HORASSEMANA * 52/12

modelv.lm<-lm(modelo, data=Varones)
summary(modelv.lm)
```

```
##
## Call:
## lm(formula = modelo, data = Varones)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1799.8  -375.1  -75.4   271.7  5988.0
##
## Coefficients: (1 not defined because of singularities)
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    160.99644    59.57558   2.702  0.00691 **
## ESUPERIORES     792.31934    28.32578  27.972 < 2e-16 ***
## ESECUNDARIOS    235.56088    25.88801   9.099 < 2e-16 ***
## EPRIMARIOS         NA         NA      NA      NA
## EXPERIENCIA     27.82960     2.86501   9.714 < 2e-16 ***
## I(EXPERIENCIA^2) -0.29505     0.05937  -4.970 6.91e-07 ***
## ENFERMEDAD     -52.64739    21.34541  -2.466  0.01368 *
## HORASMENSUALES   2.89480     0.27119  10.674 < 2e-16 ***
## SUPERVISOR      378.95922    19.19615  19.741 < 2e-16 ***
## TEMPORAL       -185.17982    22.31282  -8.299 < 2e-16 ***
## EXTRANJERO     -193.39174    30.32692  -6.377 1.95e-10 ***
## CASADO         141.60032    23.19455   6.105 1.10e-09 ***
## Numerohijos12   27.45788    12.38663   2.217  0.02668 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 602.1 on 5518 degrees of freedom
## Multiple R-squared:  0.3939, Adjusted R-squared:  0.3926
## F-statistic: 325.9 on 11 and 5518 DF, p-value: < 2.2e-16
```

```
testestbr<-efp(modelo, type="Rec-CUSUM", data = Varones, h=0.2)
plot(testestbr)
```

### Recursive CUSUM test





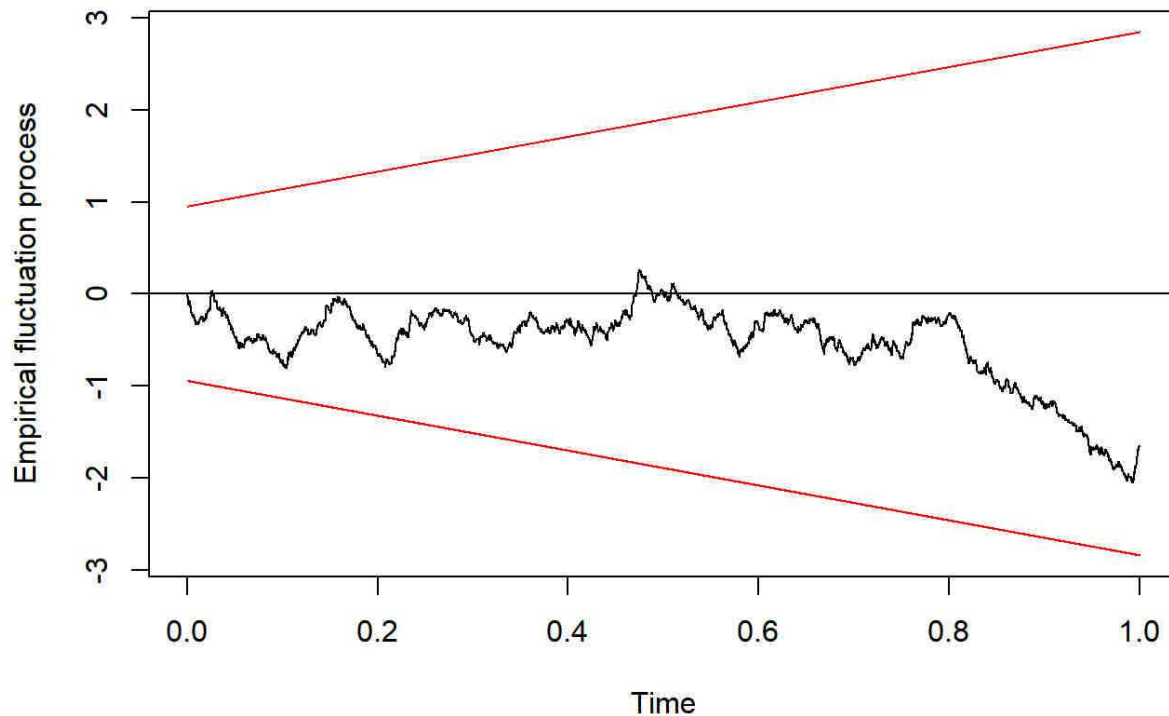
```
#MODELO MUJERES:
Mujeres$HORASMENSUALES <- Mujeres$HORASSEMANA * 52/12

modelm.lm<-lm(modelo, data=Mujeres)
summary(modelm.lm)
```

```
##
## Call:
## lm(formula = modelo, data = Mujeres)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2042.9  -318.4   -47.3    257.9  4841.1
##
## Coefficients: (1 not defined because of singularities)
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)   -360.74585    44.54714  -8.098 6.98e-16 ***
## ESUPERIORES     819.70409    27.44606   29.866 < 2e-16 ***
## ESECUNDARIOS    196.14176    26.99755    7.265 4.31e-13 ***
## EPRIMARIOS           NA          NA         NA      NA
## EXPERIENCIA     24.58094     2.53878    9.682 < 2e-16 ***
## I(EXPERIENCIA^2)  -0.19293     0.06028   -3.201 0.00138 **
## ENFERMEDAD      -22.72154    18.91619   -1.201 0.22974
## HORASMENSUALES     5.35419     0.19027   28.139 < 2e-16 ***
## SUPERVISOR       244.18910    19.78396   12.343 < 2e-16 ***
## TEMPORAL        -78.62618    18.70432   -4.204 2.67e-05 ***
## EXTRANJERO      -165.07827    25.56651   -6.457 1.17e-10 ***
## CASADO           65.58552    16.74763    3.916 9.12e-05 ***
## Numerohijos12     20.82226    10.87687    1.914 0.05563 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 510.7 on 4927 degrees of freedom
## Multiple R-squared:  0.4787, Adjusted R-squared:  0.4775
## F-statistic: 411.3 on 11 and 4927 DF,  p-value: < 2.2e-16
```

```
testestbr<-efp(modelo, type="Rec-CUSUM", data = Mujeres, h=0.2)
plot(testestbr)
```

## Recursive CUSUM test



```
# Entre las mujeres tampoco hay cambio estructural

datos$salarioEcVarones <- predict(modelv.lm, newdata = datos)
```

```
## Warning in predict.lm(modelv.lm, newdata = datos): prediction from a rank-
## deficient fit may be misleading
```

```
datos$salarioEcMujeres <- predict(modelm.lm, newdata = datos)
```

```
## Warning in predict.lm(modelm.lm, newdata = datos): prediction from a rank-
## deficient fit may be misleading
```

```
aa<-datos%>%
  group_by(MUJER)%>%
  summarise(mediaSalarioecV=mean(salarioEcVarones,na.rm = TRUE), mediaSalarioecM=mean(salarioEcMujeres, na.rm=TRUE))
aa
```

```
## # A tibble: 2 x 3
##   MUJER mediaSalarioecV mediaSalarioecM
##   <dbl>         <dbl>         <dbl>
## 1     0         1645.         1474.
## 2     1         1536.         1334.
```

```
#write.table(aa, "clipboard", sep=";", row.names=FALSE)
```

```
#Salarios medios por sexo
tapply(datos$SALARIO, datos$MUJER, mean) #summary
```

```
##           0           1
## 1645.131 1333.577
```

```
bb<-datos%>%
  group_by(MUJER)%>%
  summarise(mediaSalario=mean(SALARIO,na.rm = TRUE))
bb
```

```
## # A tibble: 2 x 2
##   MUJER mediaSalario
##   <dbl>         <dbl>
## 1     0         1645.
## 2     1         1334.
```

```
# write.table(bb, "clipboard", sep=";", row.names=FALSE)

# Indicadores de discriminación
# ¿Cuanto tendría que subir el salario de la mujer para ser remunerada como si fuera un varón?
aa[2,2]/aa[2,3]-1
```

```
##   mediaSalarioecV
## 1           0.152152
```

```
# ¿Cuanto tendría que bajar el salario de los varones si fuesen remunerada como si fueran mujeres?
aa[1,3]/aa[1,2]-1
```

```
##   mediaSalarioecM
## 1          -0.104217
```

```
# ¿Cual es la brecha salarial?
bb[1,2]-bb[2,2]
```

```
##   mediaSalario
## 1       311.5538
```

```
aa[1,2]-aa[2,3]
```

```
##   mediaSalarioecV
## 1           311.5538
```

```
# ¿Que parte de la brecha salarial se debe a diferencia de características?
aa[1,2]-aa[2,2]
```

```
##   mediaSalarioecV
## 1           108.6474
```

```
(aa[1,2]-aa[2,2])/(aa[1,2]-aa[2,3])
```

```
##   mediaSalarioecV
## 1           0.3487277
```

```
# ¿Que parte de la brecha salarial se debe a discriminación?  
aa[2,2]-aa[2,3]
```

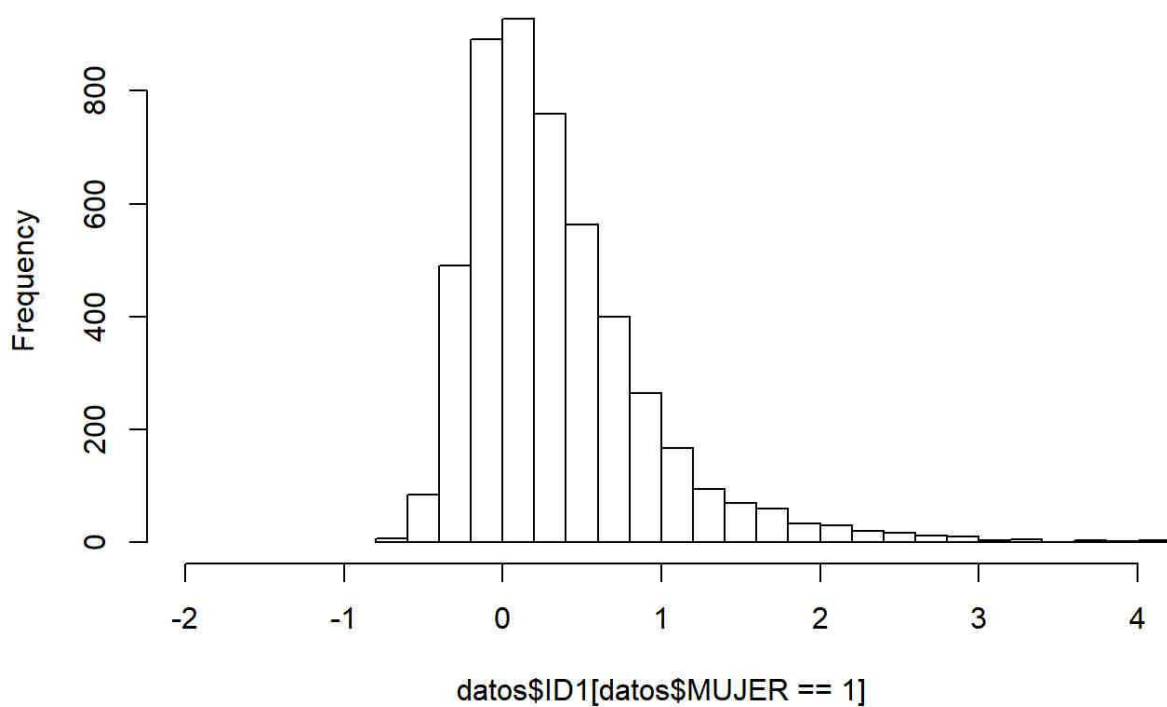
```
## mediaSalarioecV  
## 1 202.9063
```

```
(aa[2,2]-aa[2,3])/(aa[1,2]-aa[2,3])
```

```
## mediaSalarioecV  
## 1 0.6512723
```

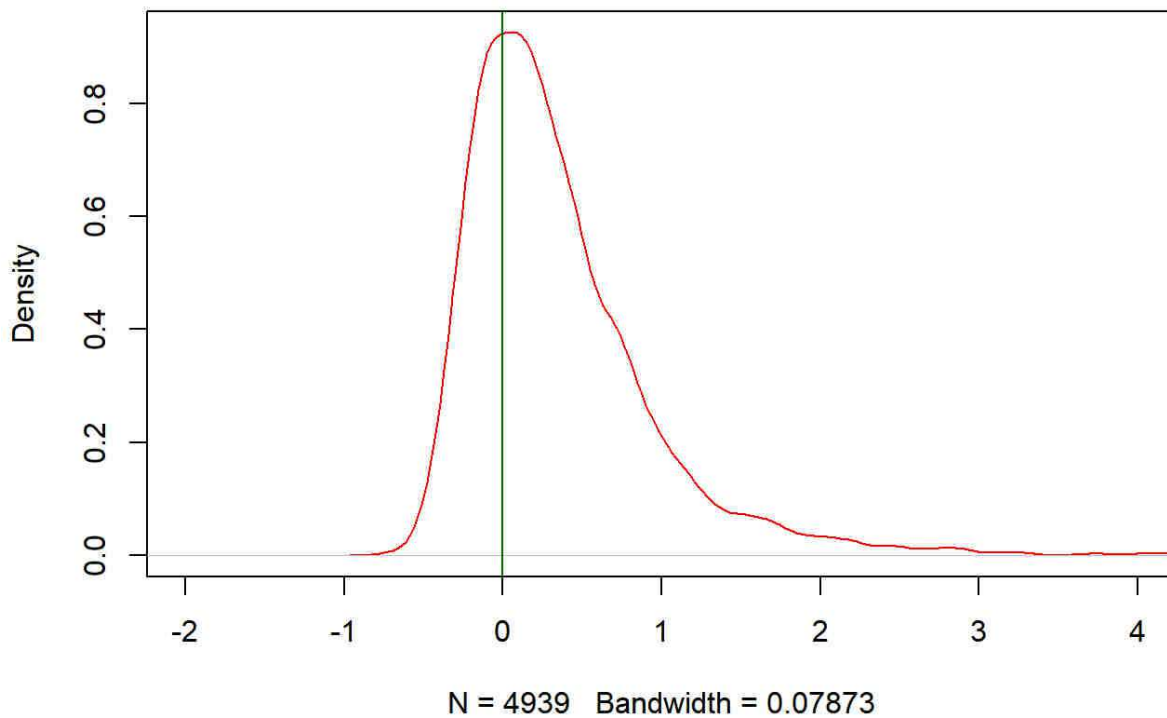
```
#Histograma de la discriminacion de las mujeres  
#¿todas las mujeres estan igual de discriminadas?  
datos$ID1 <- datos$salarioEcVarones/datos$SALARIO - 1  
  
hist(datos$ID1[datos$MUJER==1], xlim=c(-2,4), breaks = 50)
```

**Histogram of datos\$ID1[datos\$MUJER == 1]**



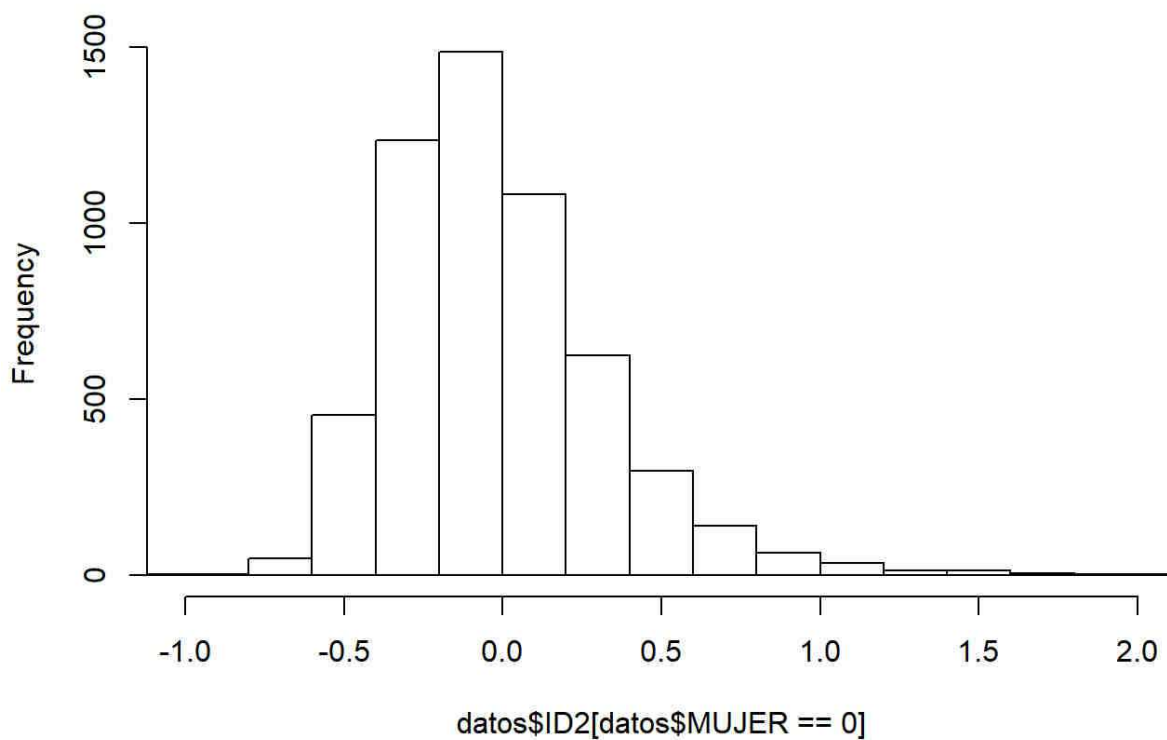
```
plot(density(datos$ID1[datos$MUJER==1]), xlim=c(-2,4), col="red")  
abline(v=0, col="darkgreen")
```

**density.default(x = datos\$ID1[datos\$MUJER == 1])**



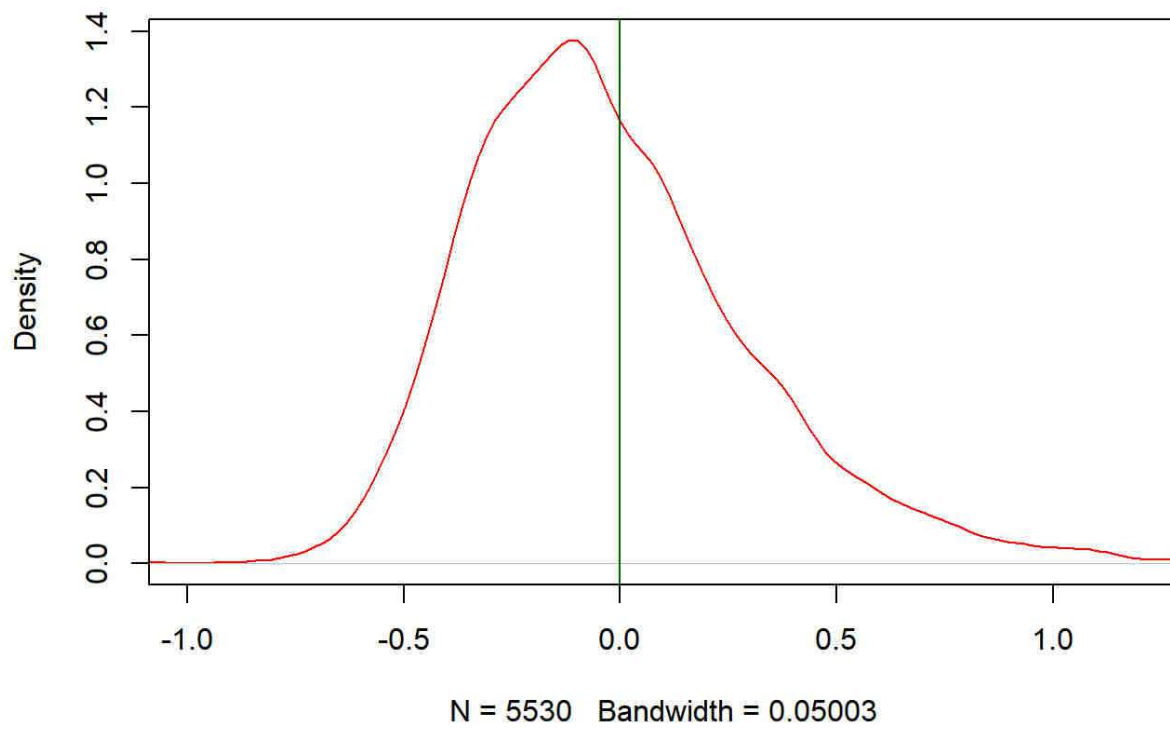
```
datos$ID2 <- datos$salarioEcMujeres/datos$SALARIO - 1  
  
hist(datos$ID2[datos$MUJER==0], xlim=c(-1,2), breaks = 50)
```

**Histogram of datos\$ID2[datos\$MUJER == 0]**



```
plot(density(datos$ID2[datos$MUJER==0]), xlim=c(-1,1.2), col="red")  
abline(v=0, col="darkgreen")
```

**density.default(x = datos\$ID2[datos\$MUJER == 0])**





Proyecto de innovación  
“Innova Docencia” de la UCM  
Convocatoria 2018-2019  
Proyecto número 275

**R-adaptación de la asignatura de métodos econométricos en economía y finanzas del  
Grado de Estadística Aplicada**

Facultad de Estudios Estadísticos  
Departamento de Economía Aplicada, Pública y Política

**Título**

***Práctica 6 : Detección y estimación robusta ante heterocedasticidad: Un ejemplo sencillo con el gasto medio por persona en las CC.AA.***

**Correspondiente al**

**Capítulo 3 MODELOS CON DATOS DE SECCIÓN CRUZADA I:  
HETEROCEDASTICIDAD**

- Causas de la heterocedasticidad
- Contrastes de heterocedasticidad
- El Estimador de Mínimos Cuadrados Generalizados
- El Estimador de Mínimos Cuadrados ponderados factible

***Práctica 6 : Detección y estimación robusta ante heterocedasticidad: Un ejemplo sencillo con el gasto medio por persona en las CC.AA.***

***Práctica 7 : Heterocedasticidad: La demanda de Cigarrillos ¿Qué política antitabaco es más eficiente?***

# Detección y estimación robusta ante Heterocedasticidad: un ejemplo

*Asignatura*

*Métodos Econométricos en Economía y Finanzas*

*Profesores*

*Loren Escot, Alicia Pérez-Alonso y Julio E. Sandubete*

Dpto Economía Aplicada, Pública y Política

Facultad Estudios Estadísticos - Universidad Complutense de Madrid

*Curso académico 2018/2019*

## **Abstract**

En esta práctica se muestra cómo ejecutar en R el ejemplo de los apuntes sobre la presencia de heterocedasticidad en el término de error de un modelo de regresión.

Se mostrarán diferentes test para contrastar la existencia de heterocedasticidad así como la estimación robusta de la matriz de varianzas y covarianzas MCO, y también la estimación por mínimos cuadrados ponderados.

- 1 Carga de librerías y datos para la práctica
- 2 Primer modelo: MCO sin considerar heterocedasticidad
  - 2.1 Test de heteroscedasticidad
- 3 ¿Qué hacer cuando se detecta Heterocedasticidad en los residuos MCO?
- 4 Estimación por mínimos cuadrados ponderados para ganar precisión
  - 4.1 ¿Hemos corregido el problema de heteroscedasticidad?
- 5 Otra forma de corregir la heteroscedasticidad: Transformación logarítmica
  - 5.1 ¿Entonces cual especificación es mejor?

## 1 Carga de librerías y datos para la práctica

```
library(stargazer) # para sacar tablas de resultados
```

```
## Warning: package 'stargazer' was built under R version 3.4.1
```

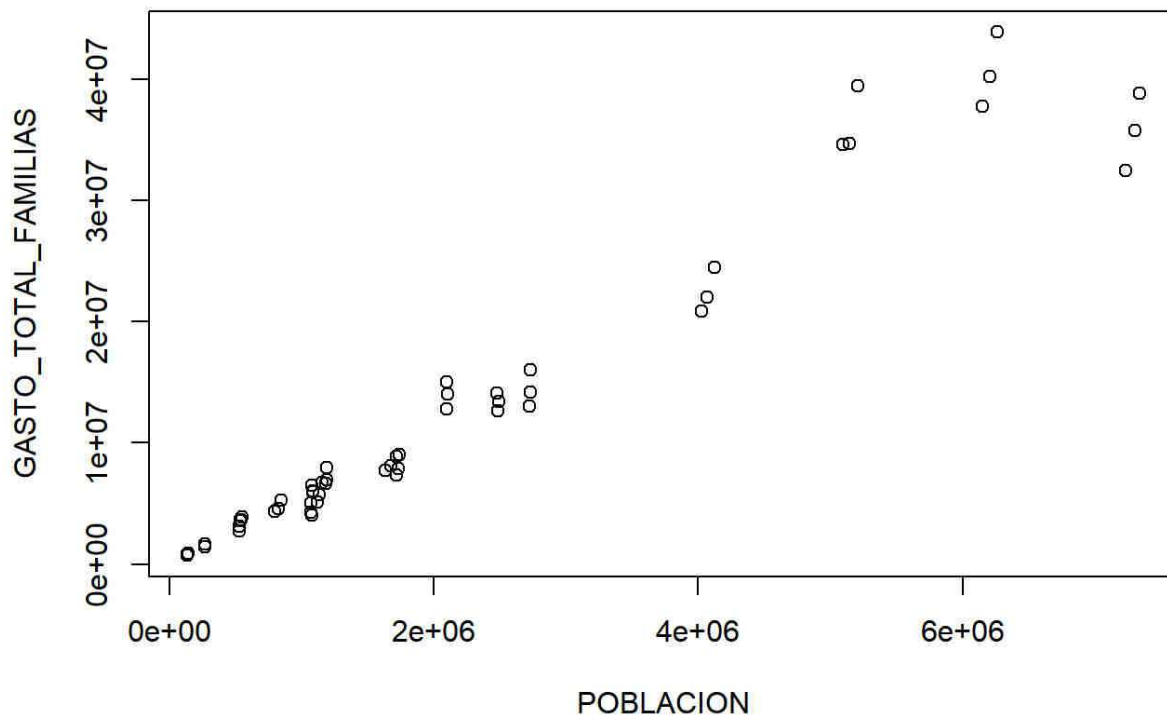
```
##  
## Please cite as:
```

```
## Hlavac, Marek (2015). stargazer: Well-Formatted Regression and Summary Statistics Tables.
```

```
## R package version 5.2. http://CRAN.R-project.org/package=stargazer
```

```
# cargo los datos utilizados en los apuntes (que relacionan población y gasto  
# total de las familias en las diferentes comunidades autónomas en diferentes años  
  
datos<-read.csv("datos_heteroscedasticidad_ejemploApuntes.csv", sep=";")  
  
plot(GASTO_TOTAL_FAMILIAS ~ POBLACION, data=datos)
```





```
#text(x = datos$POBLACION, y= datos$GASTO_TOTAL_FAMILIAS, labels = datos$CCAA)
```

## 2 Primer modelo: MCO sin considerar heterocedasticidad

```
modelo.ls<-lm(GASTO_TOTAL_FAMILIAS ~ POBLACION, data=datos) # initial model
summary(modelo.ls)
```

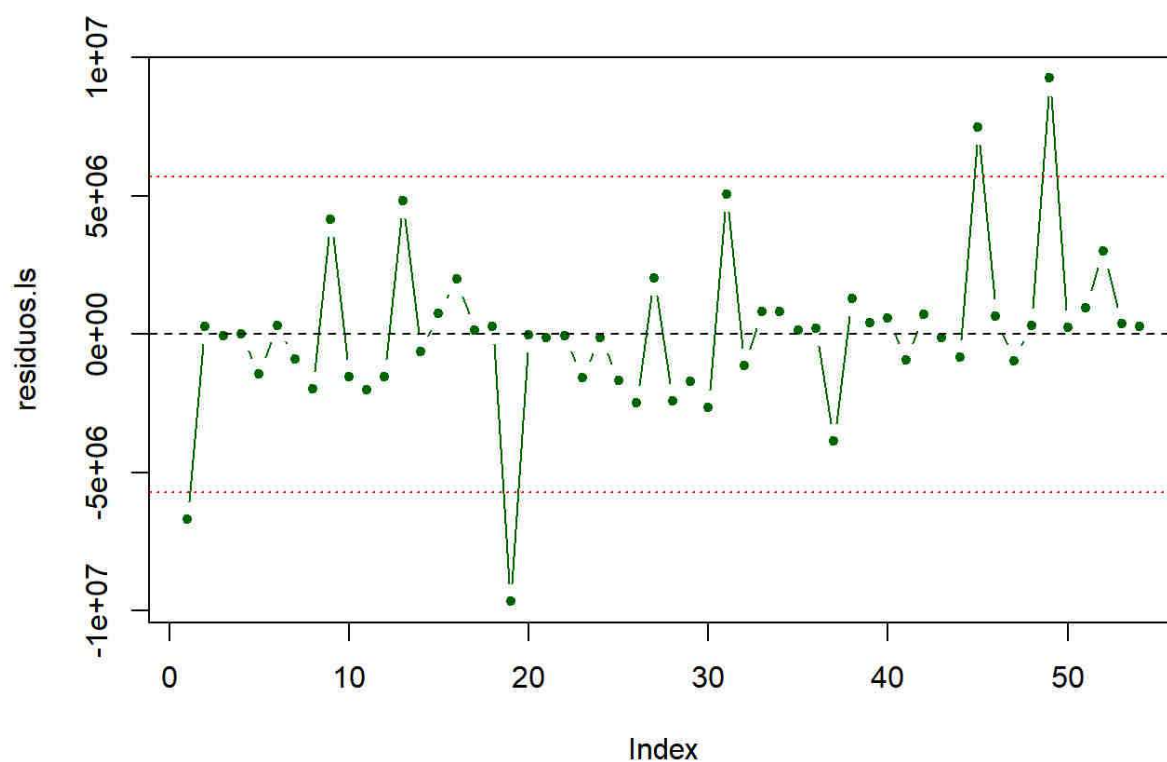
```
##
## Call:
## lm(formula = GASTO_TOTAL_FAMILIAS ~ POBLACION, data = datos)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -9665903 -1377302    70968   689680  9254658
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -2.550e+05  5.805e+05  -0.439   0.662
## POBLACION    5.854e+00  1.918e-01  30.521 <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2881000 on 52 degrees of freedom
## Multiple R-squared:  0.9471, Adjusted R-squared:  0.9461
## F-statistic: 931.5 on 1 and 52 DF, p-value: < 2.2e-16
```



```
# diagnosis de los residuos

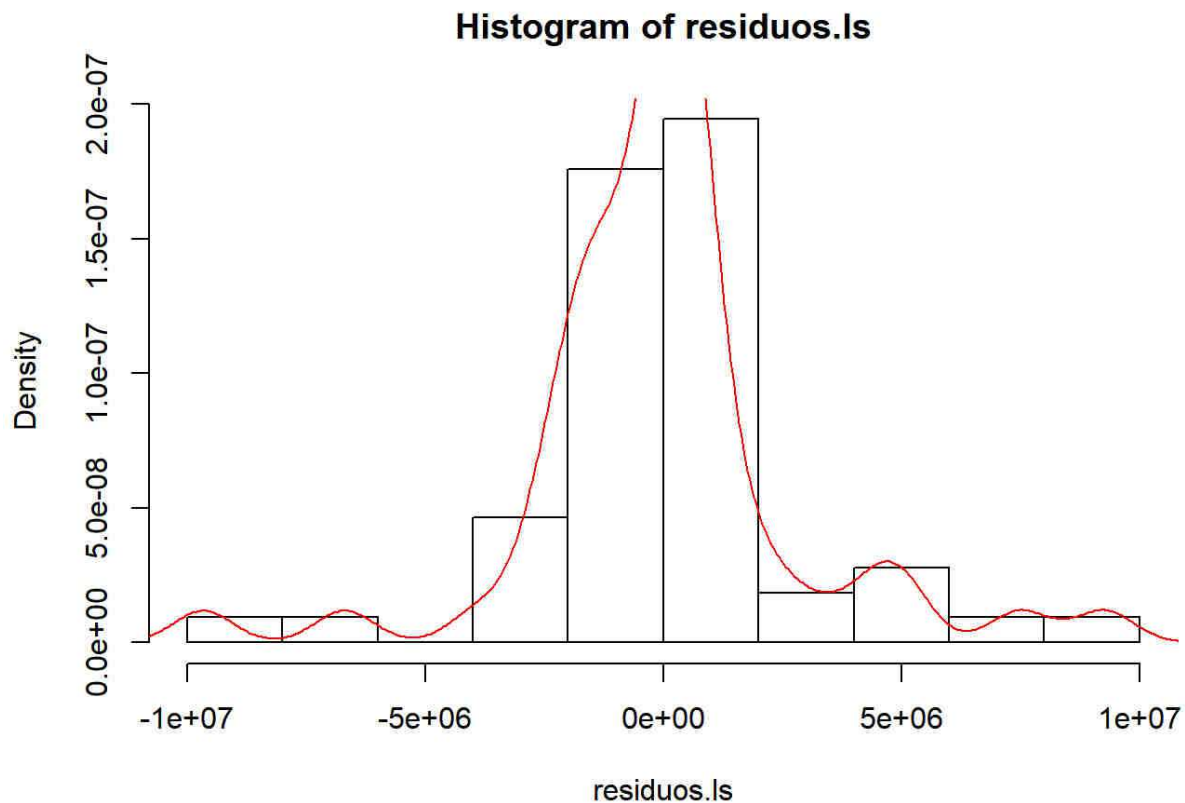
residuos.ls<-residuals(modelo.ls)

plot(residuos.ls, type="b", col="darkgreen",pch=20)
abline(h=0,lty=2, col="grey2" )
abline(h=2*sd(residuos.ls),lty=3, col="red" )
abline(h=-2*sd(residuos.ls),lty=3, col="red" )
```



```
hist(residuos.ls, freq = FALSE)
lines(density(residuos.ls),col=2)

library(tseries)
```



```
jarque.bera.test(residuos.ls) # Ho: Normalidad
```

```
##  
## Jarque Bera Test  
##  
## data:  residuos.ls  
## X-squared = 31.739, df = 2, p-value = 1.282e-07
```

```
#performs the Shapiro-Wilk test for normality  
shapiro.test(residuos.ls) # Ho: Normalidad
```

```
##  
## Shapiro-Wilk normality test  
##  
## data:  residuos.ls  
## W = 0.87385, p-value = 3.996e-05
```

## 2.1 Test de heteroscedasticidad

```
# Test de heteroscedasticidad  
library(lmtest)
```

```
## Warning: package 'lmtest' was built under R version 3.4.2
```

```
## Loading required package: zoo
```

```
##  
## Attaching package: 'zoo'
```

```
## The following objects are masked from 'package:base':
##
##   as.Date, as.Date.numeric
```

```
lmtest::bptest(modelo.ls, studentize=TRUE, data = datos) # Breusch-Pagan test H0: Ausencia de Heteroscedasticidad, usa las variables originales
```

```
##
## studentized Breusch-Pagan test
##
## data: modelo.ls
## BP = 26.07, df = 1, p-value = 3.293e-07
```

```
# Es equivalente a
lmtest::bptest(formula = modelo.ls, varformula = ~ POBLACION, studentize=TRUE, data = datos) # Breusch-Pagan test H0: Ausencia de Heteroscedasticidad, usa las variables originales
```

```
##
## studentized Breusch-Pagan test
##
## data: modelo.ls
## BP = 26.07, df = 1, p-value = 3.293e-07
```

```
# Test de White
lmtest::bptest(modelo.ls, ~ POBLACION + I(POBLACION^2), data = datos, studentize = TRUE) # Test de white poniendo cuadrados y cruzados
```

```
##
## studentized Breusch-Pagan test
##
## data: modelo.ls
## BP = 28.669, df = 2, p-value = 5.953e-07
```

```
# Manualmente
modelo.BP<-lm(residuals(modelo.ls,type = )^2 ~ POBLACION, data=datos)
summary(modelo.BP)
```

```
##
## Call:
## lm(formula = residuals(modelo.ls, type = )^2 ~ POBLACION, data = datos)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2.942e+13 -3.960e+12 -3.485e+11  3.570e+12  5.831e+13
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -6.537e+12  2.827e+12  -2.312   0.0248 *
## POBLACION    6.508e+06  9.341e+05   6.967 5.59e-09 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.403e+13 on 52 degrees of freedom
## Multiple R-squared:  0.4828, Adjusted R-squared:  0.4728
## F-statistic: 48.54 on 1 and 52 DF, p-value: 5.591e-09
```

```
tbp<-summary(modelo.BP)$r.squared*(modelo.BP$df.residual+length(modelo.BP$coefficients))
print(tbp)
```

```
## [1] 26.0697
```

```
1-pchisq(tbp,length(modelo.BP$coefficients)-1)
```

```
## [1] 3.293113e-07
```

```
modelo.WH<-lm(modelo.ls$residuals^2 ~ POBLACION + I(POBLACION^2), data=datos)
summary(modelo.WH)
```

```
##
## Call:
## lm(formula = modelo.ls$residuals^2 ~ POBLACION + I(POBLACION^2),
##     data = datos)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -3.632e+13 -1.046e+12 -2.253e+11  7.466e+11  6.135e+13
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)   4.486e+11  4.089e+12   0.110   0.9131
## POBLACION     -1.105e+06  3.447e+06  -0.321   0.7498
## I(POBLACION^2) 1.093e+00  4.777e-01   2.287   0.0264 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.349e+13 on 51 degrees of freedom
## Multiple R-squared:  0.5309, Adjusted R-squared:  0.5125
## F-statistic: 28.86 on 2 and 51 DF,  p-value: 4.143e-09
```

```
tbp<-summary(modelo.WH)$r.squared*(modelo.WH$df.residual+length(modelo.WH$coefficients))
print(tbp)
```

```
## [1] 28.66853
```

```
1-pchisq(tbp,length(modelo.WH$coefficients)-1)
```

```
## [1] 5.9526e-07
```

```
# Otra forma de hacer el test de Breusch-Pagan, pero mejor porque sirve para modelos
# lineales y modelos ponderados
```

```
library(car)
car::ncvTest(modelo.ls) # Es el test de Breusch-Pagan, suponiendo por defecto
```

```
## Non-constant Variance Score Test
## Variance formula: ~ fitted.values
## Chisquare = 74.80303    Df = 1    p = 5.200995e-18
```

```
# que la varianza del término de error depende del nivel
# de los propios valores ajustados (o de otro conjunto de
# combinaciones lineales de las variables predictoras)
```

```
car::ncvTest(modelo.ls,~ POBLACION) # test de Breusch-Pagan varianza depende de los regresores
```

```
## Non-constant Variance Score Test
## Variance formula: ~ POBLACION
## Chisquare = 74.80303    Df = 1    p = 5.200995e-18
```

```
car::ncvTest(modelo.ls,~ POBLACION + I(POBLACION^2)) # test de White la varianza depende de los
```

```
## Non-constant Variance Score Test
## Variance formula: ~ POBLACION + I(POBLACION^2)
## Chisquare = 82.25998    Df = 2    p = 1.372373e-18
```

```
# sus cuadrados y productos cruzados
```

```
ncvTest(modelo.ls) # Breusch-Pagan test H0: Ausencia de Heteroscedasticidad, varianza depende de los
propios valores ajustados
```

```
## Non-constant Variance Score Test
## Variance formula: ~ fitted.values
## Chisquare = 74.80303    Df = 1    p = 5.200995e-18
```

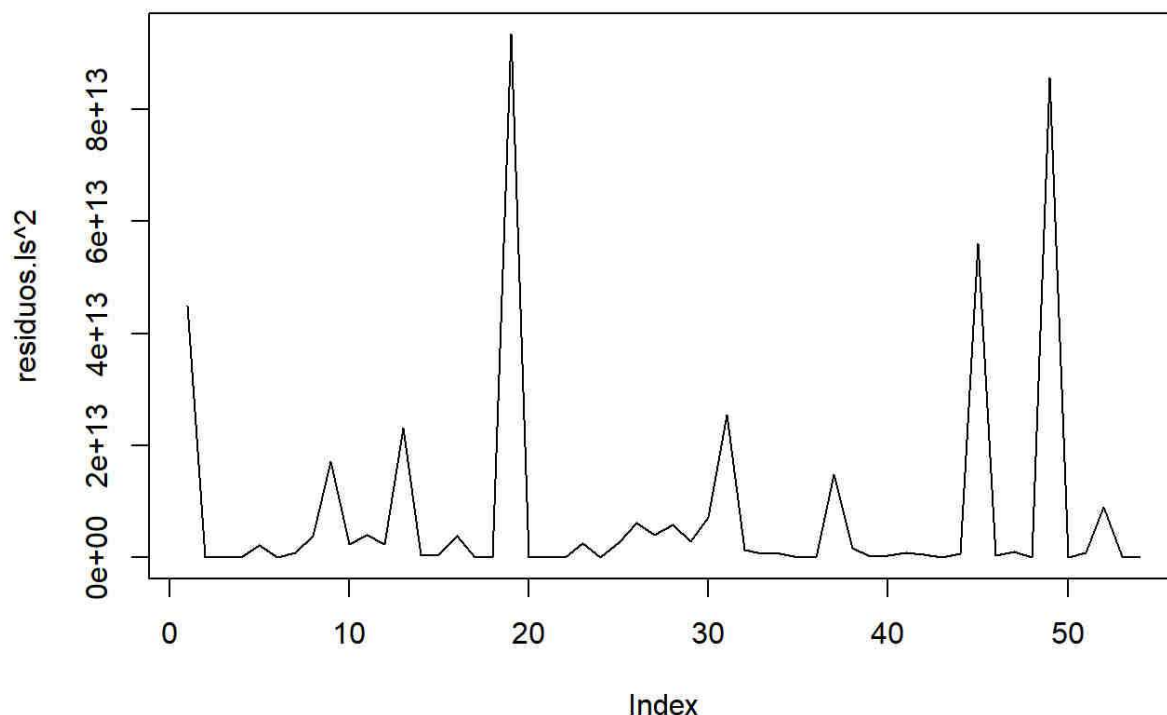
```
ncvTest(modelo.ls, ~ POBLACION) # Breusch-Pagan test H0: Ausencia de Heteroscedasticidad, usa las v
ariables originales
```

```
## Non-constant Variance Score Test
## Variance formula: ~ POBLACION
## Chisquare = 74.80303    Df = 1    p = 5.200995e-18
```

```
ncvTest(modelo.ls, ~ POBLACION + I(POBLACION^2)) # Test de white
```

```
## Non-constant Variance Score Test
## Variance formula: ~ POBLACION + I(POBLACION^2)
## Chisquare = 82.25998    Df = 2    p = 1.372373e-18
```

```
plot(residuos.ls^2, type="l")
```



### 3 ¿Qué hacer cuando se detecta Heterocedasticidad en los residuos MCO?

Dado que existe un problema evidente de Heteroscedasticidad:

1.- Estimaremos consistentemente la verdadera matriz de varianzas y covarianzas de los coeficientes estimados por MCO 2.- Aplicaremos Mínimos cuadrados ponderados para ver si podemos mejorar la precisión (menor varianza de los coeficientes estimados) 3.- Podemos probar con la transformación logarítmica (ojo con la interpretación de los coeficientes) # Estimación robusta de la verdadera matriz de varianzas y Covarianzas MCO ante presencia de Heterocedasticidad

```
# 1.- Estimación robusta de la matriz de varianzas y Covarianzas ----
# Dos alternativas
# Primera
library(car)
car::hccm(modelo.ls, type=c("hc0")) # Esta es una estimación robusta de la matriz
```

```
##           (Intercept)      POBLACION
## (Intercept)  2.315762e+11 -1.579523e+05
## POBLACION   -1.579523e+05  1.247337e-01
```

# De la matriz de Varianzas y Covarianzas

```
#se puede probar con diferentes estimaciones de la matriz de Varainzas y Covarianzas
car::hccm(modelo.ls, type=c("hc1"))# robust; HC1 (Stata default)
```

```
##           (Intercept)      POBLACION
## (Intercept)  2.404830e+11 -1.640274e+05
## POBLACION   -1.640274e+05  1.295311e-01
```



```
# Segunda (me gusta más)
library(sandwich)
sandwich::vcovHC(modelo.ls, type=c("HC0"))
```

```
##              (Intercept)      POBLACION
## (Intercept)  2.315762e+11 -1.579523e+05
## POBLACION   -1.579523e+05  1.247337e-01
```

```
# y hay varias formas de aproximar la verdadera matriz de varcov de beta MCO
#"hc0": corrección de White
#"hc1", "hc2", and "hc3" corrections are described in Long and Ervin (2000);
#"hc4" is described in Cribari-Neto (2004).
```

```
sandwich::vcovHC(modelo.ls, "HC1")# robust; HC1 (Stata default)
```

```
##              (Intercept)      POBLACION
## (Intercept)  2.404830e+11 -1.640274e+05
## POBLACION   -1.640274e+05  1.295311e-01
```

```
#

# Para sacar una tabla de resultados de la regresión con la nueva estimación
# utilizamos la libreria lmtest

# sustituir la estimación robusta en los test de significatividad
lmtest::coeftest(modelo.ls, vcov = hccm(modelo.ls, type=c("hc0")))# robust librería car; HC0
```

```
##
## t test of coefficients:
##
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -2.5497e+05  4.8122e+05 -0.5298    0.5985
## POBLACION    5.8537e+00  3.5318e-01 16.5744 <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
lmtest::coeftest(modelo.ls, vcov = sandwich)# robust; sandwich
```

```
##
## t test of coefficients:
##
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -2.5497e+05  4.8122e+05 -0.5298    0.5985
## POBLACION    5.8537e+00  3.5318e-01 16.5744 <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
lmtest::coeftest(modelo.ls, vcov = vcovHC(modelo.ls, "HC0")) # robust sandwich ; HC0
```

```
##
## t test of coefficients:
##
##           Estimate Std. Error t value Pr(>|t|)
## (Intercept) -2.5497e+05 4.8122e+05 -0.5298  0.5985
## POBLACION    5.8537e+00 3.5318e-01 16.5744 <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
modelo.lsrob<-coeftest(modelo.ls, vcov = vcovHC(modelo.ls, "HC0")) # robust sandwich ; HC0

stargazer(modelo.ls, modelo.lsrob, type="text" )
```

```
##
## =====
##                               Dependent variable:
##                               -----
##                               GASTO_TOTAL_FAMILIAS
##                               OLS                coefficient
##                               test
##                               (1)                (2)
## -----
## POBLACION                5.854***                5.854***
##                               (0.192)                (0.353)
##
## Constant                -254,969.400                -254,969.400
##                               (580,536.800)                (481,223.700)
##
## -----
## Observations                54
## R2                0.947
## Adjusted R2                0.946
## Residual Std. Error 2,880,838.000 (df = 52)
## F Statistic                931.534*** (df = 1; 52)
## =====
## Note:                *p<0.1; **p<0.05; ***p<0.01
```

```
#También el tes F de signifiación Conjunta se puede estimar con vcov robusta
lmtest::waldtest(modelo.ls,vcov = hccm(modelo.ls, type=c("hc1")))
```

```
## Wald test
##
## Model 1: GASTO_TOTAL_FAMILIAS ~ POBLACION
## Model 2: GASTO_TOTAL_FAMILIAS ~ 1
##   Res.Df Df    F    Pr(>F)
## 1      52
## 2      53 -1 264.54 < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
lmtest::waldtest(modelo.ls,vcov = vcovHC(modelo.ls, "HC0"))
```

```
## Wald test
##
## Model 1: GASTO_TOTAL_FAMILIAS ~ POBLACION
## Model 2: GASTO_TOTAL_FAMILIAS ~ 1
##   Res.Df Df       F    Pr(>F)
## 1      52
## 2      53 -1 274.71 < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

## 4 Estimación por mínimos cuadrados ponderados para ganar precisión

```
# 2.- Estimación por mínimos cuadrados ponderados para ganar precisión ----

#PARA ESTIMAR POR MINIMOS CUADRADOS PONDERADOS
# mínimos cuadrados ponderados minimiza la suma de cuadrados de los residuos
# ponderados por w: SUM(w*e^2)
# Por eso para estimar por mínimos cuadrados generalizados o mínimos cuadrados ponderados
# o bien premultiplicamos por 1/sd o bien utilizamos la ponderación w=1/var.
# El resultado es equivalente

datos$w = 1/datos$POBLACION^2

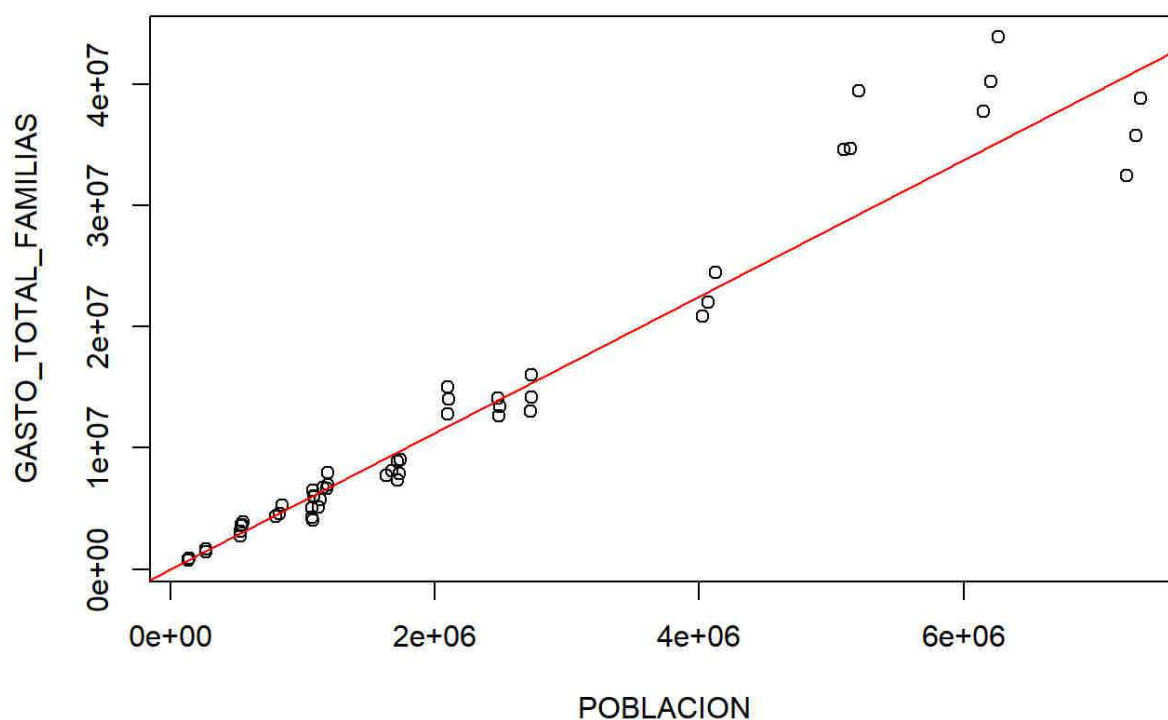
modelo.pls<-lm(GASTO_TOTAL_FAMILIAS ~ POBLACION, data=datos, weights = w) # initial model
summary(modelo.pls)
```

```
##
## Call:
## lm(formula = GASTO_TOTAL_FAMILIAS ~ POBLACION, data = datos,
##     weights = w)
##
## Weighted Residuals:
##      Min        1Q    Median        3Q       Max
## -1.9092 -0.5416 -0.1089  0.5134  1.9508
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 3.981e+04  6.826e+04   0.583   0.562
## POBLACION    5.624e+00  1.463e-01  38.449 <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.8644 on 52 degrees of freedom
## Multiple R-squared:  0.966, Adjusted R-squared:  0.9654
## F-statistic: 1478 on 1 and 52 DF, p-value: < 2.2e-16
```

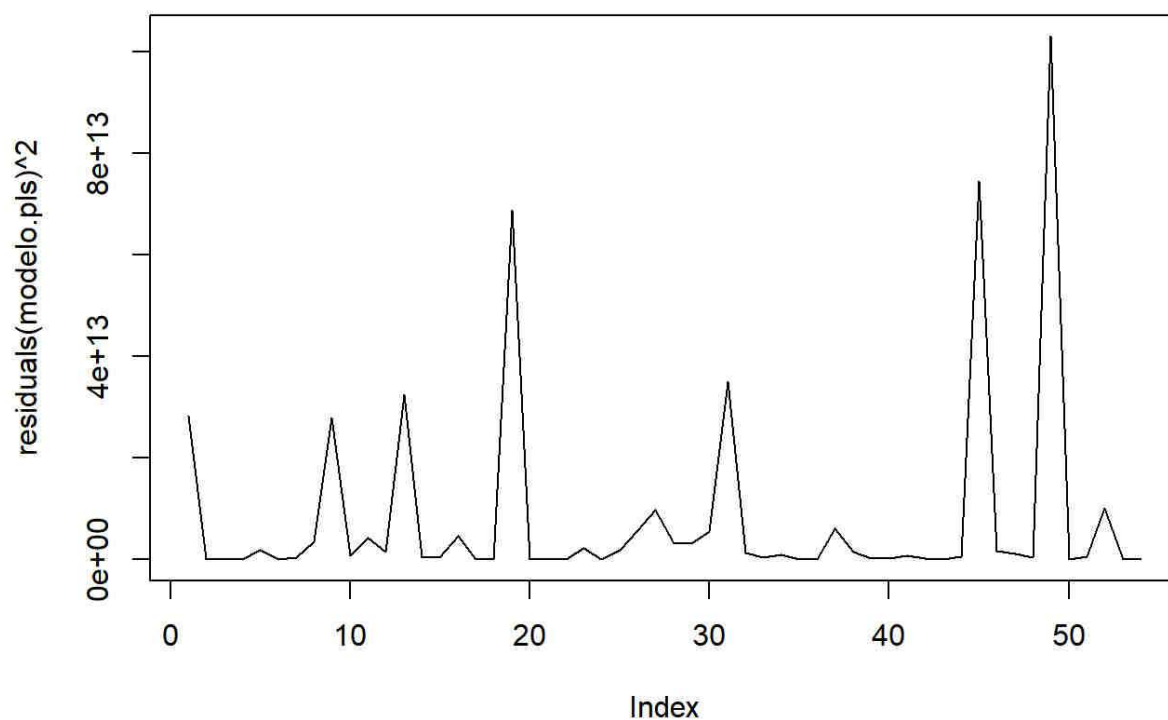
```
stargazer(modelo.lsrob, modelo.pls, type="text")
```

```
##
## =====
##                               Dependent variable:
##                               -----
##                               GASTO_TOTAL_FAMILIAS
##                               coefficient      OLS
##                               test
##                               (1)            (2)
## -----
## POBLACION          5.854***            5.624***
##
##                   (0.353)            (0.146)
##
## Constant          -254,969.400        39,813.540
##                   (481,223.700)      (68,258.850)
##
## -----
## Observations                54
## R2                        0.966
## Adjusted R2                0.965
## Residual Std. Error        0.864 (df = 52)
## F Statistic                1,478.362*** (df = 1; 52)
## =====
## Note:                    *p<0.1; **p<0.05; ***p<0.01
```

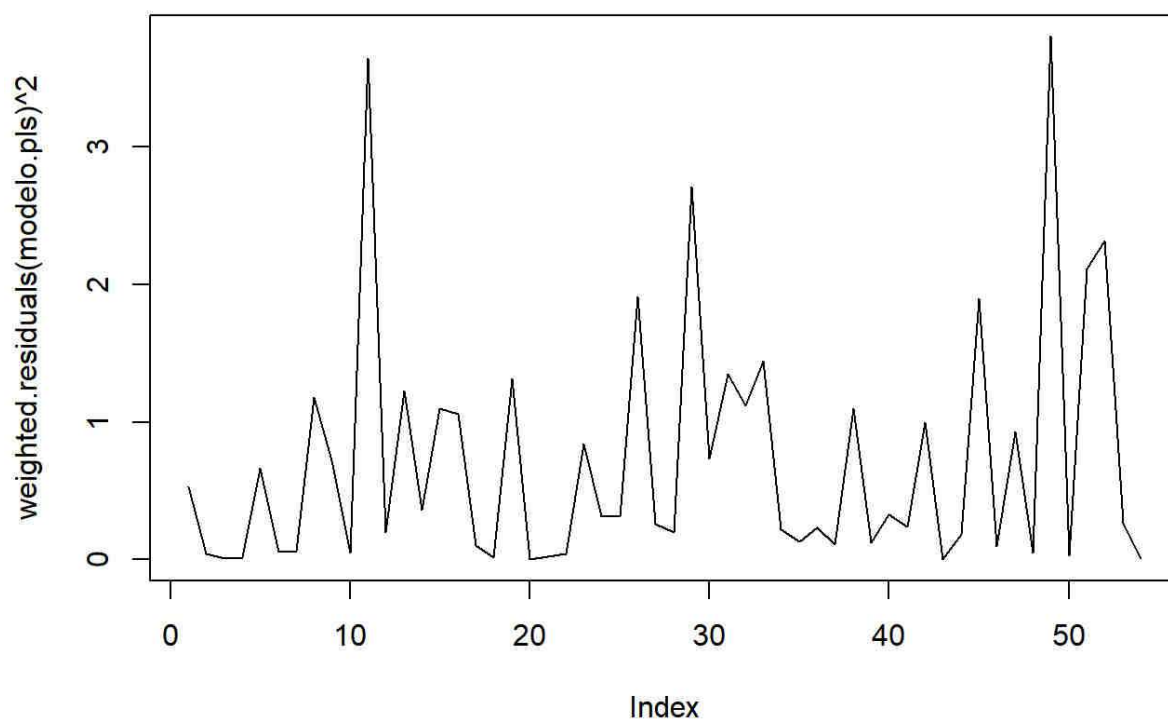
```
plot(GASTO_TOTAL_FAMILIAS ~ POBLACION, data=datos)
abline(modelo.pls, col=2)
```



```
plot(residuals(modelo.pls)^2, type="l")
```



```
plot(weighted.residuals(modelo.pls)^2, type="l")
```



4.1 ¿Hemos corregido el problema de heteroscedasticidad?

```
# ¿hemos corregido el problema de heteroscedasticidad?
datos$residuos.pls<-weighted.residuals(modelo.pls)

auxhetreg<-lm(residuos.pls^2~ POBLACION+ I(POBLACION^2), data=datos)
summary(auxhetreg)
```

```
##
## Call:
## lm(formula = residuos.pls^2 ~ POBLACION + I(POBLACION^2), data = datos)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.8554 -0.6081 -0.3458  0.4035  3.0004
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)   4.954e-01  2.747e-01   1.804  0.0772 .
## POBLACION     1.529e-07  2.316e-07   0.660  0.5121
## I(POBLACION^2) -1.279e-14  3.209e-14  -0.399  0.6918
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.9064 on 51 degrees of freedom
## Multiple R-squared:  0.0244, Adjusted R-squared:  -0.01386
## F-statistic: 0.6377 on 2 and 51 DF,  p-value: 0.5327
```

```
tbp<-summary(auxhetreg)$r.squared*(auxhetreg$df.residual+length(auxhetreg$coefficients))
print(tbp)
```

```
## [1] 1.317499
```

```
1-pchisq(tbp,length(auxhetreg$coefficients)-1)
```

```
## [1] 0.5174982
```

```
bptest(modelo.pls, data = datos,studentize = FALSE) # Breusch-Pagan test H0: Ausencia de Heterosced
asticidad, usa las variables originales
```

```
##
## Breusch-Pagan test
##
## data:  modelo.pls
## BP = 74.803, df = 1, p-value < 2.2e-16
```

```
bptest(modelo.pls, ~ POBLACION + I(POBLACION^2), data = datos) # Test de white poniendo cuadrados y
cruzados
```

```
##
## studentized Breusch-Pagan test
##
## data:  modelo.pls
## BP = 28.669, df = 2, p-value = 5.953e-07
```

```
# OJO el test lmtest::bptest falla porque no tiene en cuenta los residuos ponderados,
# que son los que habría que utilizar, así que mejor utilizar car::ncvTest

ncvTest(modelo.pls) # Breusch-Pagan test H0: Ausencia de Heteroscedasticidad, varianza depende de los
# propios valores ajustados
```

```
## Non-constant Variance Score Test
## Variance formula: ~ fitted.values
## Chisquare = 0.8858528    Df = 1    p = 0.3466036
```

```
ncvTest(modelo.pls, ~ POBLACION) # Breusch-Pagan test H0: Ausencia de Heteroscedasticidad, usa las
# variables originales
```

```
## Non-constant Variance Score Test
## Variance formula: ~ POBLACION
## Chisquare = 0.8858528    Df = 1    p = 0.3466036
```

```
ncvTest(modelo.pls, ~ POBLACION + I(POBLACION^2)) # Test de white poniendo cuadrados y cruzados
```

```
## Non-constant Variance Score Test
## Variance formula: ~ POBLACION + I(POBLACION^2)
## Chisquare = 1.011986    Df = 2    p = 0.6029067
```

## 5 Otra forma de corregir la heteroscedasticidad: Transformación logarítmica

```
# Otra forma de corregir la heteroscedasticidad: Transformación logarítmica ?????

modelo.log<-lm(log(GASTO_TOTAL_FAMILIAS) ~ log(POBLACION), data=datos) # initial model
summary(modelo.log)
```

```
##
## Call:
## lm(formula = log(GASTO_TOTAL_FAMILIAS) ~ log(POBLACION), data = datos)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.40338 -0.08948 -0.01098  0.10267  0.30662
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.78230     0.28957   6.155 1.09e-07 ***
## log(POBLACION)  0.99592     0.02039  48.836 < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.1556 on 52 degrees of freedom
## Multiple R-squared:  0.9787, Adjusted R-squared:  0.9783
## F-statistic: 2385 on 1 and 52 DF, p-value: < 2.2e-16
```

```
bptest(modelo.log, data = datos ) # Breusch-Pagan test H0: Ausencia de Heteroscedasticidad, usa las
# variables originales
```

```
##  
## studentized Breusch-Pagan test  
##  
## data: modelo.log  
## BP = 1.0075, df = 1, p-value = 0.3155
```

```
bptest(modelo.log, ~ I(log(POBLACION)^2), data = datos) # Test de white poniendo cuadrados y cruzados
```

```
##  
## studentized Breusch-Pagan test  
##  
## data: modelo.log  
## BP = 0.92485, df = 1, p-value = 0.3362
```

```
ncvTest(modelo.log) # Breusch-Pagan test H0: Ausencia de Heteroscedasticidad, varianza depende de los propios valores ajustados
```

```
## Non-constant Variance Score Test  
## Variance formula: ~ fitted.values  
## Chisquare = 0.9025135 Df = 1 p = 0.3421086
```

```
ncvTest(modelo.log, ~ log(POBLACION)) # Breusch-Pagan test H0: Ausencia de Heteroscedasticidad, usa las variables originales
```

```
## Non-constant Variance Score Test  
## Variance formula: ~ log(POBLACION)  
## Chisquare = 0.9025135 Df = 1 p = 0.3421086
```

```
ncvTest(modelo.log, ~ log(POBLACION) + I(log(POBLACION)^2)) # Test de white
```

```
## Non-constant Variance Score Test  
## Variance formula: ~ log(POBLACION) + I(log(POBLACION)^2)  
## Chisquare = 1.60674 Df = 2 p = 0.4478173
```

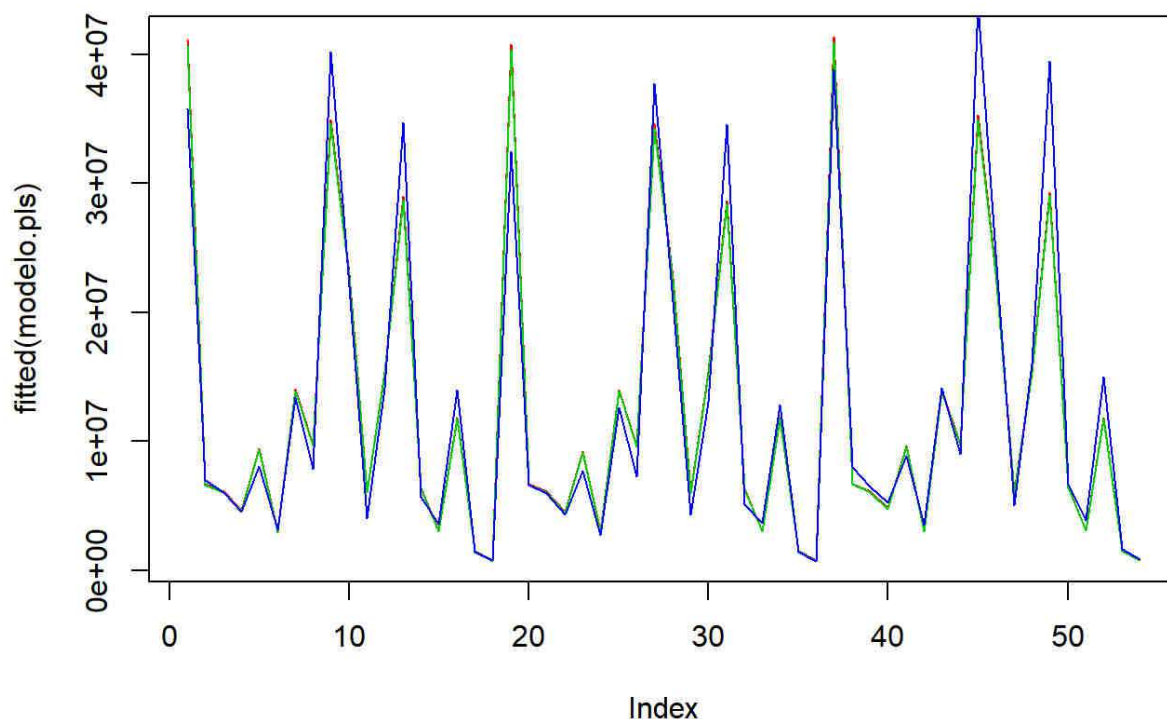
## 5.1 ¿Entonces cual especificación es mejor?

```
# ¿Entonces cual especificación es mejor?  
stargazer(modelo.pls, modelo.log, type="text")
```

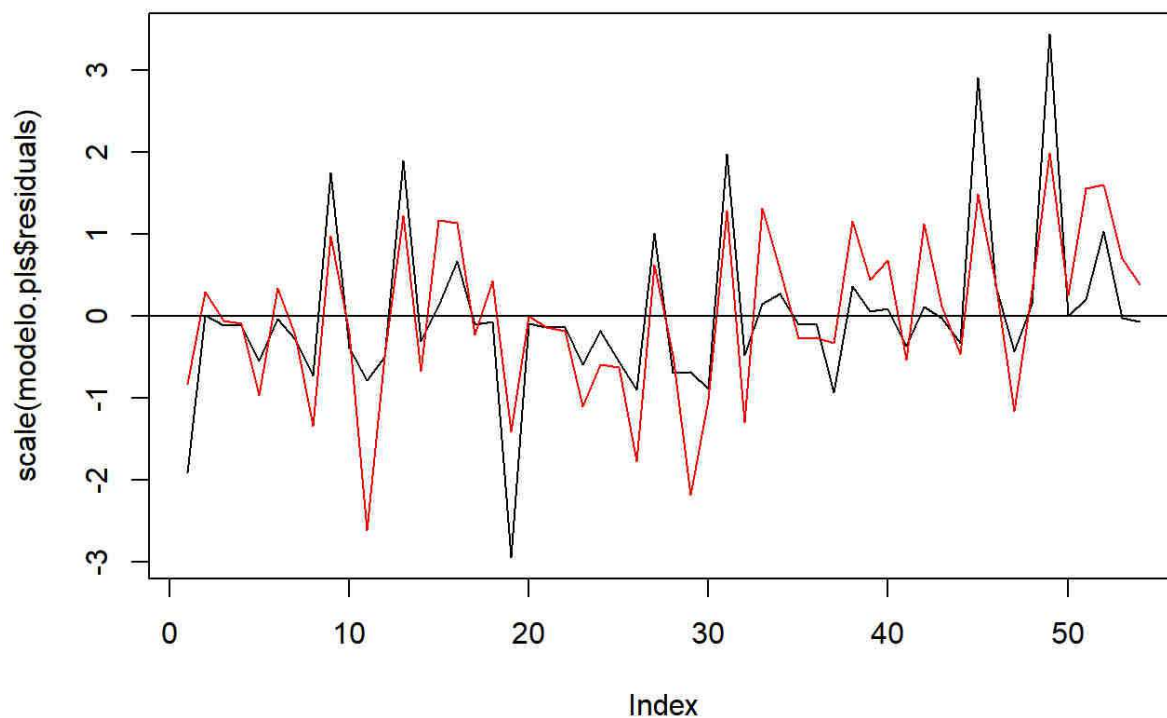


```
##
## =====
##                               Dependent variable:
##                               -----
##                               GASTO_TOTAL_FAMILIAS log(GASTO_TOTAL_FAMILIAS)
##                               (1)                   (2)
##                               -----
## POBLACION                    5.624***
##                               (0.146)
##
## log(POBLACION)                                0.996***
##                                               (0.020)
##
## Constant                    39,813.540          1.782***
##                               (68,258.850)        (0.290)
##
## -----
## Observations                54                  54
## R2                          0.966                0.979
## Adjusted R2                 0.965                0.978
## Residual Std. Error (df = 52) 0.864                0.156
## F Statistic (df = 1; 52)      1,478.362***        2,384.918***
## =====
## Note:                        *p<0.1; **p<0.05; ***p<0.01
```

```
plot(fitted(modelo.pls), col=2, type="l")
lines(exp(fitted(modelo.log)), col=3, type="l")
lines(modelo.pls$model[,1],col="blue", type="l")
```



```
plot(scale(modelo.pls$residuals), type="l")
lines(scale(modelo.log$residuals), type="l", col=2)
abline(h=0)
```



```
# Contraste de Especificación de Ramsey (Ho: el modelo está bien especificado)
# Test RESET añade como variables adicionales yhat^2 + yhat^3 + yhat^4....
# y luego hace un test de significación conjunta de esos coeficientes
# la idea es detectar algún tipo de función polinómica general
# para la mala especificación
```

```
library(lmtest)
resettest(modelo.pls) # (Ho: el modelo está bien especificado)
```

```
##
## RESET test
##
## data: modelo.pls
## RESET = 22.762, df1 = 2, df2 = 50, p-value = 9.364e-08
```

```
resettest(modelo.log) # (Ho: el modelo está bien especificado)
```

```
##
## RESET test
##
## data: modelo.log
## RESET = 1.4777, df1 = 2, df2 = 50, p-value = 0.238
```

```
# Conclusión: según el test de Ramsey (RESET) la mejor especificación
# será el modelo en logs-logs
```



Proyecto de innovación  
“Innova Docencia” de la UCM  
Convocatoria 2018-2019  
Proyecto número 275

**R-adaptación de la asignatura de métodos econométricos en economía y finanzas del  
Grado de Estadística Aplicada**

Facultad de Estudios Estadísticos  
Departamento de Economía Aplicada, Pública y Política

**Título**

***Práctica 7 : Heterocedasticidad: La demanda de Cigarrillos ¿Qué política antitabaco es más eficiente?***

**Correspondiente al**

**Capítulo 3 MODELOS CON DATOS DE SECCIÓN CRUZADA I:  
HETEROCEDASTICIDAD**

- Causas de la heterocedasticidad
- Contrastes de heterocedasticidad
- El Estimador de Mínimos Cuadrados Generalizados
- El Estimador de Mínimos Cuadrados ponderados factible

***Práctica 6 : Detección y estimación robusta ante heterocedasticidad: Un ejemplo sencillo con el gasto medio por persona en las CC.AA.***

***Práctica 7 : Heterocedasticidad: La demanda de Cigarrillos ¿Qué política antitabaco es más eficiente?***

# Heterocedasticidad: Demanda de Cigarrillos

## *Asignatura Métodos Econométricos en Economía y Finanzas*

*Profesores  
Loren Escot, Alicia Pérez-Alonso y Julio E. Sandubete*

Dpto Economía Aplicada, Pública y Política  
Facultad Estudios Estadísticos - Universidad Complutense de Madrid

*Curso académico 2018/2019*

### **Abstract**

En esta práctica se muestra un ejemplo donde es necesario tener en cuenta la heterocedasticidad e el término de error. Además de los test de heterocedasticidad se muestra el algoritmo de mínimos cuadrados ponderados factible de un modelo de regresión.

El ejemplo trata de estimar una función de demanda de cigarrillos para preguntarse ¿qué resulta más efectivo para reducir el tabaco, subir los impuestos sobre el tabaco, o aumentar las políticas de prohibición de fumar en lugares públicos.

- 1 Carga de librerías
- 2 Carga de datos sobre consumo de cigarrillos
- 3 Análisis exploratorio inicial
- 4 Demanda de cigarrillos por MCO
  - 4.1 Test de heterocedasticidad
- 5 Estimación robusta de la verdadera matriz de var y cov de MCO
- 6 Mínimos cuadrados generalizados: Mínimos cuadrados factibles
  - 6.1 ¿Hemos corregido la heteroscedasticidad?

```
# Demanda de Cigarrillos
# ¿qué medida resulta más eficaz para reducir el número de cigarrillos?
```

## 1 Carga de librerías

```
library(wooldridge) # para los datos
```

```
## Warning: package 'wooldridge' was built under R version 3.4.4
```

```
library(stargazer) # para mostrar los resultados
```

```
## Warning: package 'stargazer' was built under R version 3.4.1
```

```
##
## Please cite as:
```

```
## Hlavac, Marek (2015). stargazer: Well-Formatted Regression and Summary Statistics Tables.
```

```
## R package version 5.2. http://CRAN.R-project.org/package=stargazer
```

```
library(lmtest) # para los test de heteroscedasticidad
```

```
## Warning: package 'lmtest' was built under R version 3.4.2
```

```
## Loading required package: zoo
```

```
##  
## Attaching package: 'zoo'
```

```
## The following objects are masked from 'package:base':  
##  
## as.Date, as.Date.numeric
```

```
library(sandwich) # para estimación robusta de la matriz de varianzas y covarianzas de Beta MCO  
library(ggplot2)
```

```
## Warning: package 'ggplot2' was built under R version 3.4.4
```

## 2 Carga de datos sobre consumo de cigarrillos

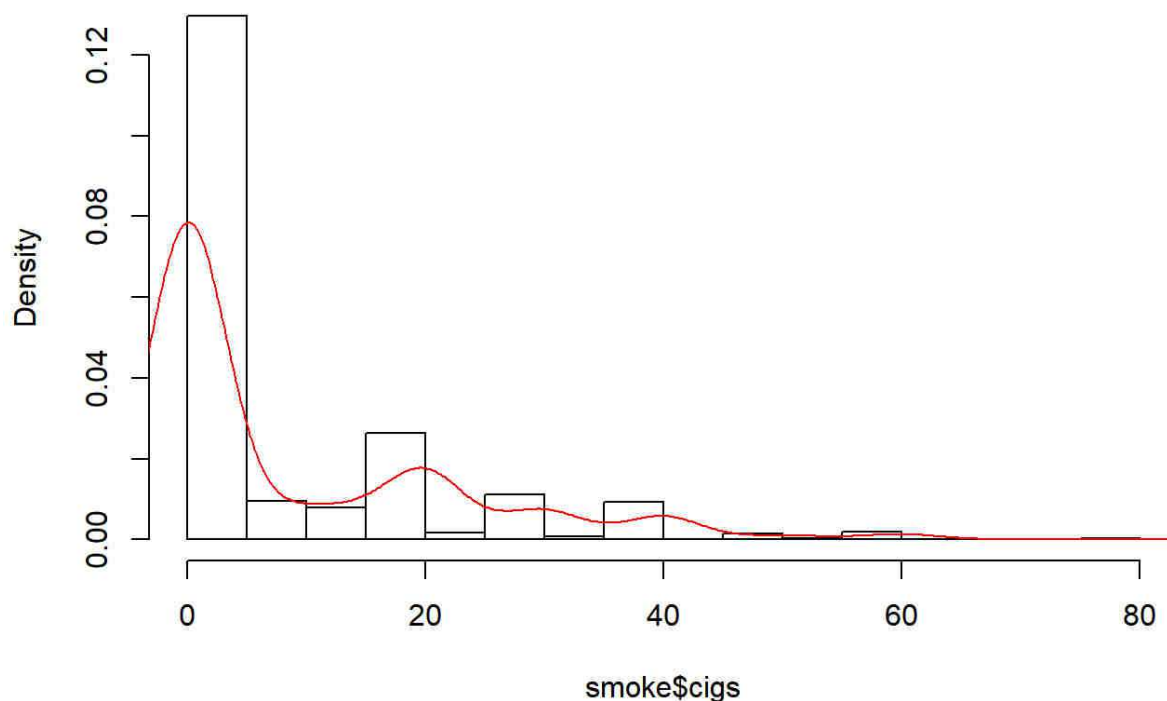
```
# es un ejemplo del libro de Wooldridge (libro de referencia básico de la asignatura)  
data('smoke')  
  
#cigs = number of cigarettes smoked per day.  
#income = annual income.  
#cigpric = the per-pack price of cigarettes (in cents).  
#educ = years of schooling.  
#age = age measured in years.  
#restaurn = a binary indicator equal to unity if the person resides in a state with restaurant smoking restrictions.  
  
#?smoke  
str(smoke)
```

```
## 'data.frame': 807 obs. of 10 variables:
## $ educ : num 16 16 12 13.5 10 6 12 15 12 12 ...
## $ cigpric : num 60.5 57.9 57.7 57.9 58.3 ...
## $ white : int 1 1 1 1 1 1 1 1 1 1 ...
## $ age : int 46 40 58 30 17 86 35 48 48 31 ...
## $ income : int 20000 30000 30000 20000 20000 6500 20000 30000 20000 20000 ...
## $ cigs : int 0 0 3 0 0 0 0 0 0 0 ...
## $ restaurn: int 0 0 0 0 0 0 0 0 0 0 ...
## $ lincome : num 9.9 10.3 10.3 9.9 9.9 ...
## $ agesq : int 2116 1600 3364 900 289 7396 1225 2304 2304 961 ...
## $ lcigpric: num 4.1 4.06 4.05 4.06 4.07 ...
## - attr(*, "datalabel")= chr ""
## - attr(*, "time.stamp")= chr "25 Jun 2011 23:03"
## - attr(*, "formats")= chr "%9.0g" "%9.0g" "%8.0g" "%8.0g" ...
## - attr(*, "types")= int 254 254 251 251 252 251 251 254 252 254
## - attr(*, "val.labels")= chr "" "" "" "" ...
## - attr(*, "var.labels")= chr "years of schooling" "state cig. price, cents/pack" "=1 if white"
"in years" ...
## - attr(*, "version")= int 10
```

### 3 Análisis exploratorio inicial

```
hist(smoke$cigs,breaks = 20, freq = FALSE)
lines(density(smoke$cigs),col=2)
```

**Histogram of smoke\$cigs**

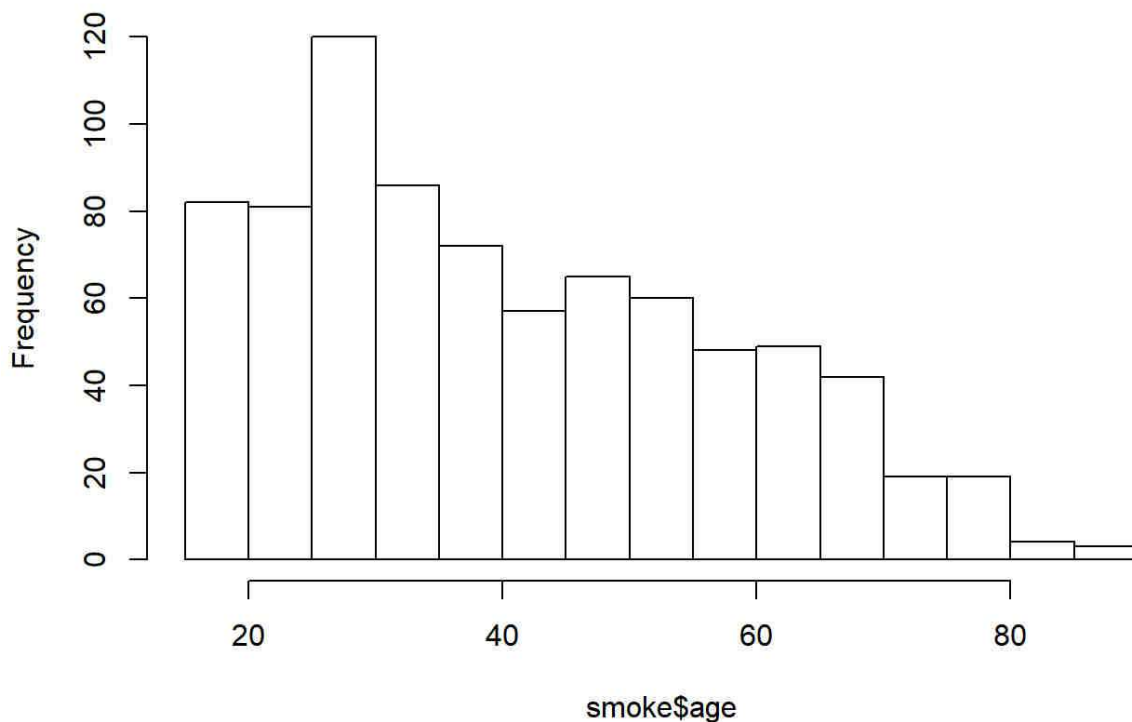


```
summary(smoke)
```

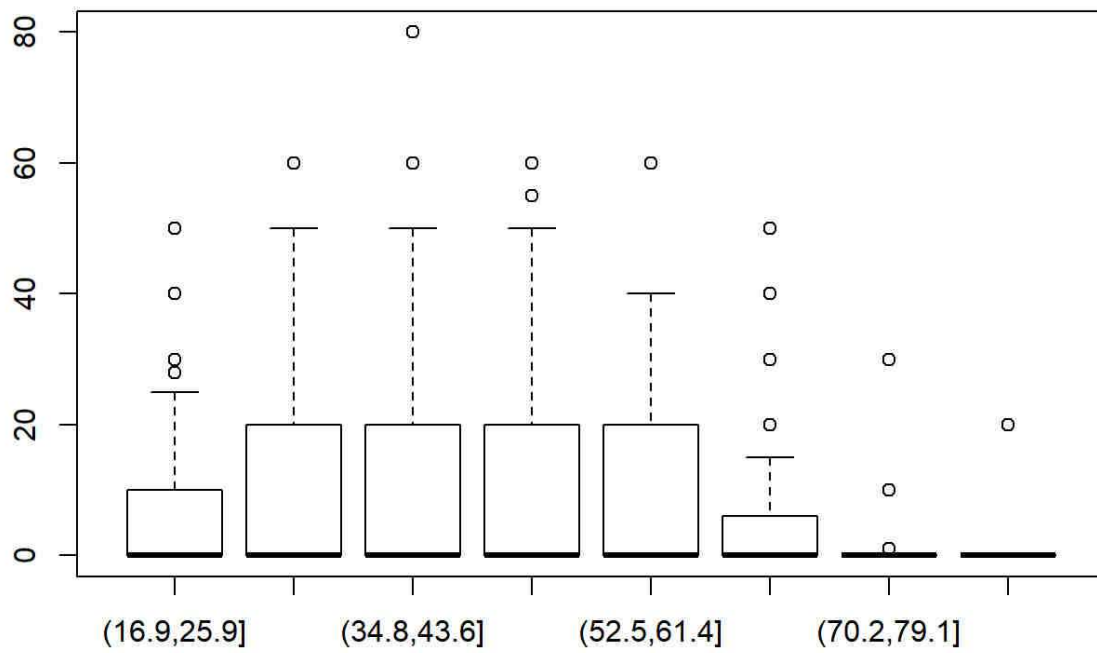
```
##      educ      cigpric      white      age
## Min.   : 6.00   Min.   :44.00   Min.   :0.0000   Min.   :17.00
## 1st Qu.:10.00   1st Qu.:58.14   1st Qu.:1.0000   1st Qu.:28.00
## Median :12.00   Median :61.05   Median :1.0000   Median :38.00
## Mean   :12.47   Mean   :60.30   Mean   :0.8786   Mean   :41.24
## 3rd Qu.:13.50   3rd Qu.:63.18   3rd Qu.:1.0000   3rd Qu.:54.00
## Max.   :18.00   Max.   :70.13   Max.   :1.0000   Max.   :88.00
##      income      cigs      restaurn      lincome
## Min.   : 500     Min.   : 0.000   Min.   :0.0000   Min.   : 6.215
## 1st Qu.:12500     1st Qu.: 0.000   1st Qu.:0.0000   1st Qu.: 9.433
## Median :20000     Median : 0.000   Median :0.0000   Median : 9.903
## Mean   :19305     Mean   : 8.686   Mean   :0.2466   Mean   : 9.687
## 3rd Qu.:30000     3rd Qu.:20.000   3rd Qu.:0.0000   3rd Qu.:10.309
## Max.   :30000     Max.   :80.000   Max.   :1.0000   Max.   :10.309
##      agesq      lcigpric
## Min.   : 289     Min.   :3.784
## 1st Qu.: 784     1st Qu.:4.063
## Median :1444     Median :4.112
## Mean   :1990     Mean   :4.096
## 3rd Qu.:2916     3rd Qu.:4.146
## Max.   :7744     Max.   :4.250
```

```
hist(smoke$age)
```

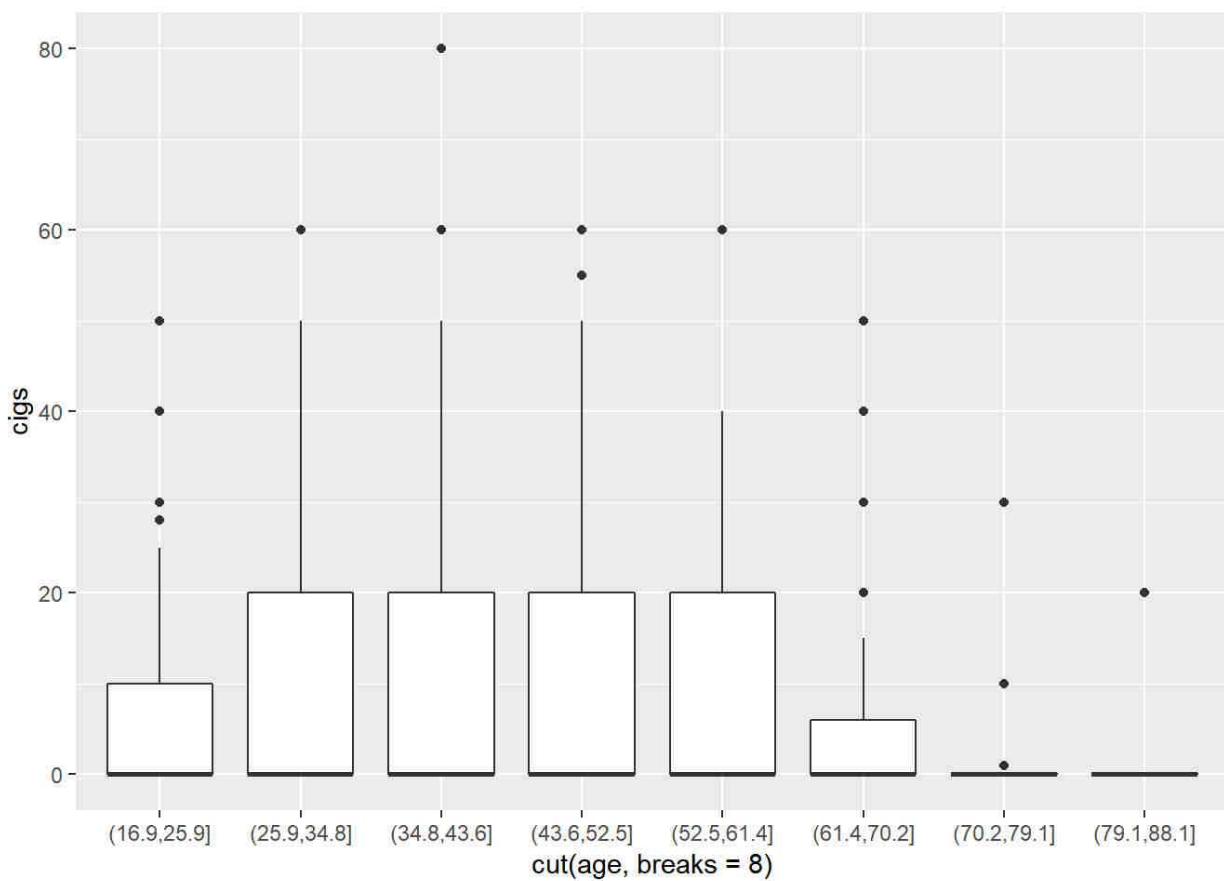
**Histogram of smoke\$age**



```
boxplot(cigs ~ cut(age,breaks = 8), data=smoke )
```



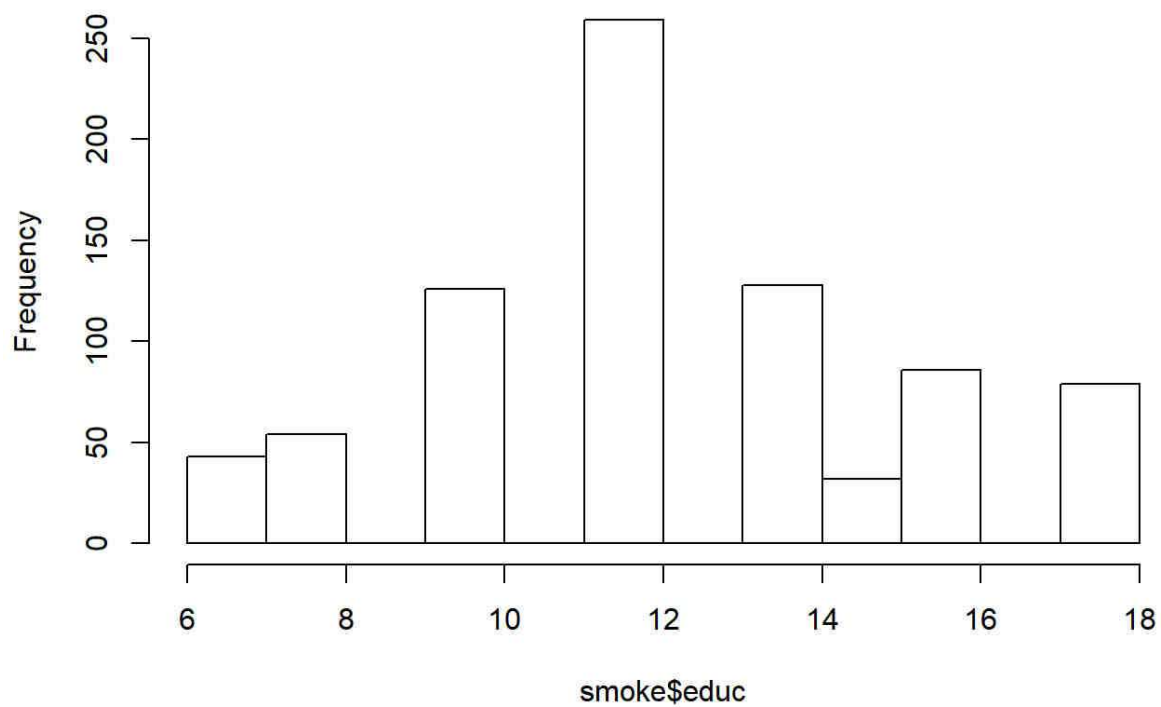
```
# el mismo gráfico con ggplot
library(ggplot2)
ggplot(data=smoke)+
  geom_boxplot(aes(y=cigs, x=cut(age,breaks = 8)))
```



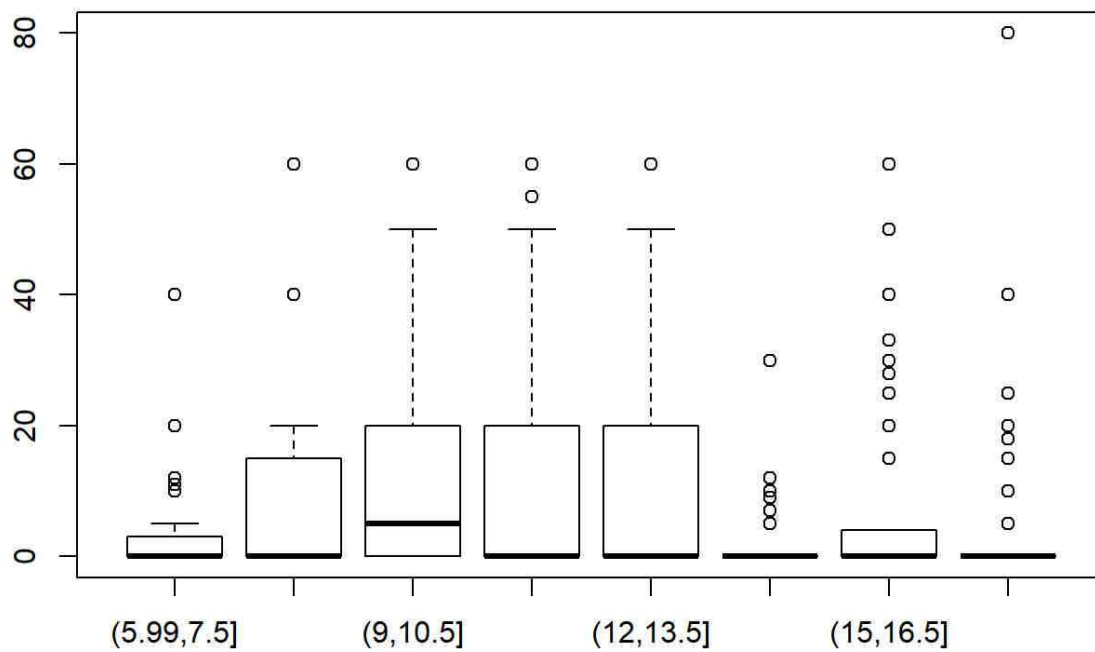
```
hist(smoke$educ)
```



**Histogram of smoke\$educ**



```
boxplot(cigs ~ cut(educ,breaks = 8), data=smoke )
```



## 4 Demanda de cigarrillos por MCO

```
# ..... conviene seguir con este tipo de análisis exploratorio
# determinantes de la demanda de cualquier bien o servicio???
# preferencias Utilidad???, y restricción presupuestaria precio y renta
# más otras restricciones

modelo.ls<-lm(cigs ~ lincome + lcigpric + educ + age + agesq + restaurn + white, data=smoke)
summary(modelo.ls)
```

```
##
## Call:
## lm(formula = cigs ~ lincome + lcigpric + educ + age + agesq +
##     restaurn + white, data = smoke)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -15.772  -9.330  -5.907   7.945  70.275
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -2.682435   24.220730  -0.111  0.91184
## lincome      0.869014    0.728764   1.192  0.23344
## lcigpric     -0.850904    5.782321  -0.147  0.88305
## educ         -0.501753    0.167168  -3.001  0.00277 **
## age          0.774502    0.160516   4.825 1.68e-06 ***
## agesq        -0.009069    0.001748  -5.188 2.70e-07 ***
## restaurn     -2.865621    1.117406  -2.565  0.01051 *
## white        -0.559236    1.459461  -0.383  0.70169
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 13.41 on 799 degrees of freedom
## Multiple R-squared:  0.05291,    Adjusted R-squared:  0.04461
## F-statistic: 6.377 on 7 and 799 DF,  p-value: 2.588e-07
```

```
stargazer(modelo.ls, type="text")
```

```
##
## =====
##               Dependent variable:
##           -----
##               cigs
## -----
## lincome           0.869
##                  (0.729)
##
## lcigpric          -0.851
##                  (5.782)
##
## educ              -0.502***
##                  (0.167)
##
## age                0.775***
##                  (0.161)
##
## agesq             -0.009***
##                  (0.002)
##
## restaurn          -2.866**
##                  (1.117)
##
## white             -0.559
##                  (1.459)
##
## Constant          -2.682
##                  (24.221)
##
## -----
## Observations           807
## R2                     0.053
## Adjusted R2            0.045
## Residual Std. Error    13.412 (df = 799)
## F Statistic            6.377*** (df = 7; 799)
## =====
## Note:                 *p<0.1; **p<0.05; ***p<0.01
```

## 4.1 Test de heterocedasticidad

Al tratarse de datos de sección cruzada sospechamos que podemos tener un posible problema de heterocedasticidad

```
# Breusch-Pagan y white
lmtest::bptest(modelo.ls,studentize = FALSE) # Breusch-Pagan test H0: Ausencia de Heteroscedasticidad, usa las variables originales
```

```
##
## Breusch-Pagan test
##
## data:  modelo.ls
## BP = 69.644, df = 7, p-value = 1.744e-12
```

```
smoke$resid2<-modelo.ls$residuals^2

modelo.BP<-lm(resid2~ lincome + lcigpric + educ + age + agesq + restaurn + white, data=smoke)
summary(modelo.BP)
```

```
##
## Call:
## lm(formula = resid2 ~ lincome + lcigpric + educ + age + agesq +
##     restaurn + white, data = smoke)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -271.7  -127.7   -93.3   -40.1  4674.3
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -661.34159   656.84588   -1.007   0.3143
## lincome      24.90689    19.76346    1.260   0.2079
## lcigpric     64.08784   156.81171    0.409   0.6829
## educ        -2.38208     4.53345   -0.525   0.5994
## age         19.39123     4.35305    4.455 9.60e-06 ***
## agesq       -0.21439     0.04741  -4.522 7.04e-06 ***
## restaurn    -70.12767    30.30311   -2.314   0.0209 *
## white       11.00069    39.57936    0.278   0.7811
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 363.7 on 799 degrees of freedom
## Multiple R-squared:  0.04012,    Adjusted R-squared:  0.03171
## F-statistic: 4.771 on 7 and 799 DF,  p-value: 2.862e-05
```

```
tbp<-summary(modelo.BP)$r.squared*(modelo.BP$df.residual+length(modelo.BP$coefficients))
print(tbp)
```

```
## [1] 32.37686
```

```
1-pchisq(tbp,length(modelo.BP$coefficients)-1)
```

```
## [1] 3.457538e-05
```

```
# TEst de White incluyendo cuadrados y productos cruzados
bptest(modelo.ls, ~ I(lincome^2)+ I(lcigpric^2)+ I(educ^2)+I(age^2)+
      I(agesq^2)+I(restaurn^2)+I(lincome*lcigpric)+I(lincome*educ)+
      I(lincome*age)+I(lincome*restaurn)+I(educ*age), data = smoke) # Test de white poniendo cuad
rados y cruzados
```

```
##
## studentized Breusch-Pagan test
##
## data: modelo.ls
## BP = 29.056, df = 11, p-value = 0.002225
```

```
modelo.BP<-lm(modelo.ls$residuals^2~ lincome + lcigpric + educ + age + agesq + restaurn + white +
      I(lincome^2)+ I(lcigpric^2)+ I(educ^2)+I(age^2)+
      I(agesq^2)+I(restaurn^2)+I(lincome*lcigpric)+I(lincome*educ)+
      I(lincome*age)+I(lincome*restaurn)+I(educ*age), data=smoke)
summary(modelo.BP)
```

```
##
## Call:
## lm(formula = modelo.ls$residuals^2 ~ lincome + lcigpric + educ +
##     age + agesq + restaurn + white + I(lincome^2) + I(lcigpric^2) +
##     I(educ^2) + I(age^2) + I(agesq^2) + I(restaurn^2) + I(lincome *
##     lcigpric) + I(lincome * educ) + I(lincome * age) + I(lincome *
##     restaurn) + I(educ * age), data = smoke)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -328.0  -132.4   -91.7   -14.9  4683.6
##
## Coefficients: (2 not defined because of singularities)
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    3.215e+04  1.959e+04   1.641  0.10121
## lincome        -1.682e+03  8.153e+02  -2.062  0.03949 *
## lcigpric       -1.244e+04  9.342e+03  -1.332  0.18320
## educ           4.769e+01  6.663e+01   0.716  0.47431
## age            4.628e+01  1.514e+01   3.057  0.00231 **
## agesq        -5.478e-01  1.905e-01  -2.875  0.00414 **
## restaurn       5.486e+02  4.385e+02   1.251  0.21127
## white          1.061e+01  3.963e+01   0.268  0.78894
## I(lincome^2)   -1.337e+00  1.617e+01  -0.083  0.93410
## I(lcigpric^2)   1.008e+03  1.155e+03   0.873  0.38319
## I(educ^2)      -2.275e-01  1.278e+00  -0.178  0.85874
## I(age^2)                NA         NA      NA      NA
## I(agesq^2)       2.142e-05  1.269e-05   1.688  0.09186 .
## I(restaurn^2)     NA         NA      NA      NA
## I(lincome * lcigpric) 4.535e+02  2.043e+02   2.220  0.02669 *
## I(lincome * educ)   -5.865e+00  7.694e+00  -0.762  0.44614
## I(lincome * age)    -1.049e+00  1.121e+00  -0.936  0.34972
## I(lincome * restaurn) -6.282e+01  4.478e+01  -1.403  0.16108
## I(educ * age)       2.568e-01  2.857e-01   0.899  0.36901
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 363.3 on 790 degrees of freedom
## Multiple R-squared:  0.0532, Adjusted R-squared:  0.03402
## F-statistic: 2.774 on 16 and 790 DF, p-value: 0.0002297
```

```
tbp<-summary(modelo.BP)$r.squared*(modelo.BP$df.residual+length(modelo.BP$coefficients))
print(tbp)
```

```
## [1] 43.03601
```

```
1-pchisq(tbp,length(modelo.BP$coefficients)-1)
```

```
## [1] 0.0007909078
```

```
library(car)
car::ncvTest(modelo.ls) # Es el test de Breusch-Pagan, suponiendo por defecto
```

```
## Non-constant Variance Score Test
## Variance formula: ~ fitted.values
## Chisquare = 55.77363    Df = 1    p = 8.131497e-14
```

```
# que la varianza del término de error depende del nivel
# de los propios valores ajustados (o de otro conjunto de
# combinaciones lineales de las variables predictoras)

car::ncvTest(modelo.ls, ~ lincome + lcigpric + educ + age + agesq + restaurn + white) # test de de
Breusch-Pagan varianza depende de los regresores
```

```
## Non-constant Variance Score Test
## Variance formula: ~ lincome + lcigpric + educ + age + agesq + restaurn + white
## Chisquare = 69.64402    Df = 7    p = 1.743539e-12
```

## 5 Estimación robusta de la verdadera matriz de var y cov de MCO

```
# Estimación robusta de la verdadera matriz de var y cov de betas MCO
modelo.lsrobust<-lmtest::coefest(modelo.ls, vcov = vcovHC(modelo.ls, "HC0")) #
print(modelo.lsrobust)
```

```
##
## t test of coefficients:
##
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -2.6824347 25.7732316 -0.1041 0.917133
## lincome      0.8690144 0.5950006 1.4605 0.144538
## lcigpric     -0.8509044 6.0243117 -0.1412 0.887712
## educ         -0.5017533 0.1616027 -3.1049 0.001971 **
## age          0.7745021 0.1373458 5.6391 2.372e-08 ***
## agesq        -0.0090686 0.0014517 -6.2469 6.799e-10 ***
## restaurn     -2.8656213 1.0122202 -2.8310 0.004756 **
## white        -0.5592363 1.3714344 -0.4078 0.683548
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
print(lmtest::waldtest(modelo.ls,vcov =vcovHC(modelo.ls, "HC0")))
```

```
## Wald test
##
## Model 1: cigs ~ lincome + lcigpric + educ + age + agesq + restaurn + white
## Model 2: cigs ~ 1
##   Res.Df Df    F   Pr(>F)
## 1     799
## 2     806 -7 9.4717 2.51e-11 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

## 6 Minimos cuadrados generalizados: Mínimos cuadrados factibles

```
#Minimos cuadrados generalizados: Mínimos cuadrados factibles
```

```
lres.u<-log(modelo.ls$residuals^2)
modelo.fgls.u<-lm(lres.u ~ lincome + lcigpric + educ + age + agesq + restaurn + white +
                  I(sqrt(age)) , data=smoke)
hhat<-exp(modelo.fgls.u$fitted.values)
modelo.fgls<-lm(cigs ~ lincome + lcigpric + educ + age + agesq + restaurn + white,
                weights=1/hhat, data=smoke)
summary(modelo.fgls)
```

```
##
## Call:
## lm(formula = cigs ~ lincome + lcigpric + educ + age + agesq +
##     restaurn + white, data = smoke, weights = 1/hhat)
##
## Weighted Residuals:
##      Min       1Q   Median       3Q      Max
## -2.4200 -0.9556 -0.7965  0.8502  9.6831
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  8.7995084  18.1129685   0.486  0.62723
## lincome      1.2801883   0.4384364   2.920  0.00360 **
## lcigpric     -3.3389722   4.5013759  -0.742  0.45845
## educ         -0.4626410   0.1214608  -3.809  0.00015 ***
## age          0.4779228   0.0980825   4.873 1.33e-06 ***
## agesq        -0.0056074   0.0009519  -5.891 5.67e-09 ***
## restaurn     -3.5568819   0.8066253  -4.410 1.18e-05 ***
## white        -1.4172531   0.9757894  -1.452  0.14678
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.572 on 799 degrees of freedom
## Multiple R-squared:  0.1142, Adjusted R-squared:  0.1064
## F-statistic: 14.71 on 7 and 799 DF,  p-value: < 2.2e-16
```

```
stargazer(modelo.ls, modelo.lsrobust, modelo.fgls,type = "text")
```

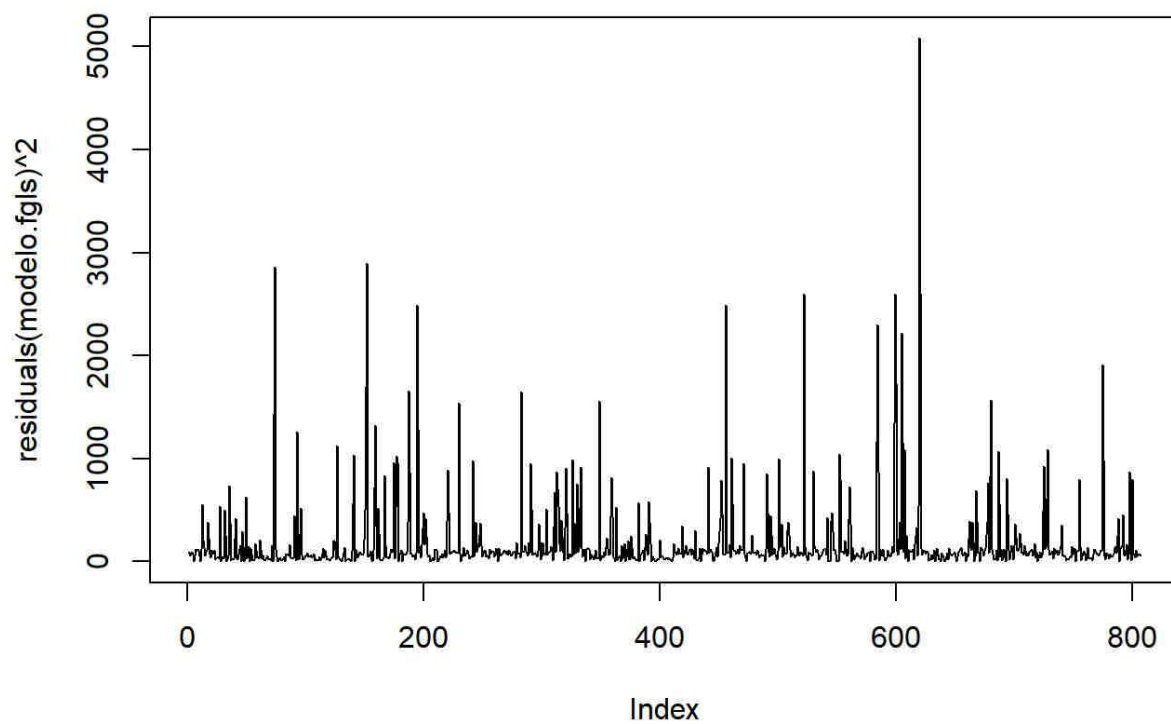
```
##
## =====
##                               Dependent variable:
##                               -----
##                               cigs      coefficient      cigs
##                               OLS      test      OLS
##                               (1)      (2)      (3)
## -----
## lincome      0.869      0.869      1.280***
##              (0.729)      (0.595)      (0.438)
##
## lcigpric     -0.851      -0.851      -3.339
##              (5.782)      (6.024)      (4.501)
##
## educ         -0.502***   -0.502***   -0.463***
##              (0.167)      (0.162)      (0.121)
##
## age          0.775***   0.775***   0.478***
##              (0.161)      (0.137)      (0.098)
##
## agesq        -0.009***   -0.009***   -0.006***
##              (0.002)      (0.001)      (0.001)
##
## restaurn     -2.866**    -2.866***   -3.557***
##              (1.117)      (1.012)      (0.807)
##
## white        -0.559      -0.559      -1.417
##              (1.459)      (1.371)      (0.976)
##
## Constant     -2.682      -2.682      8.800
##              (24.221)      (25.773)      (18.113)
## -----
## Observations      807      807
## R2                 0.053      0.114
## Adjusted R2        0.045      0.106
## Residual Std. Error (df = 799) 13.412      1.572
## F Statistic (df = 7; 799)      6.377***      14.710***
## =====
## Note:              *p<0.1; **p<0.05; ***p<0.01
```

## 6.1 ¿Hemos corregido la heteroscedasticidad?

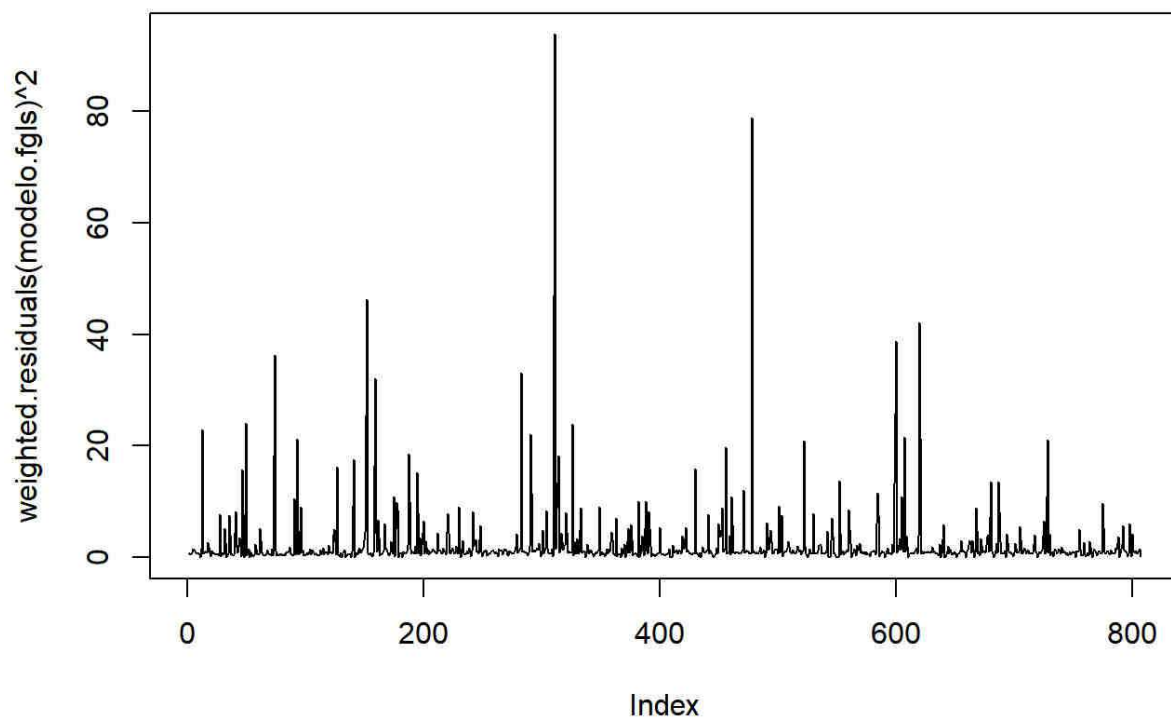
```
# ¿Hemos corregido la heteroscedasticidad?
# Ojo hay que utilizar los residuos ponderados (u_i/dt_i)

plot(residuals(modelo.fgls)^2, type="l")
```





```
plot(weighted.residuals(modelo.fgls)^2, type="l")
```



```
smoke$residuos.fgls<-weighted.residuals(modelo.fgls)
```

```
auxhetreg<-lm((residuos.fgls)^2~ lincome + lcigpric + educ + age + agesq + restaurn + white, data=smoke)
summary(auxhetreg)
```

```
##
## Call:
## lm(formula = (residuos.fgls)^2 ~ lincome + lcigpric + educ +
##     age + agesq + restaurn + white, data = smoke)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -6.979 -1.765 -1.268 -0.731  87.308
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  25.0737737  11.1767133   2.243  0.02514 *
## lincome      -0.2047399   0.3362897  -0.609  0.54282
## lcigpric     -4.3878352   2.6682659  -1.644  0.10048
## educ          0.0558355   0.0771399   0.724  0.46939
## age          -0.1748385   0.0740704  -2.360  0.01849 *
## agesq         0.0022745   0.0008066   2.820  0.00493 **
## restaurn      0.0673059   0.5156296   0.131  0.89618
## white        -0.7984274   0.6734718  -1.186  0.23616
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 6.189 on 799 degrees of freedom
## Multiple R-squared:  0.02491,    Adjusted R-squared:  0.01637
## F-statistic: 2.916 on 7 and 799 DF,  p-value: 0.005116
```

```
tbp<-summary(auxhetreg)$r.squared*(auxhetreg$df.residual+length(auxhetreg$coefficients))
print(tbp)
```

```
## [1] 20.1033
```

```
1-pchisq(tbp,length(auxhetreg$coefficients)-1)
```

```
## [1] 0.00535083
```

```
library(car)
car::ncvTest(modelo.fgls) #Computes a score test of the hypothesis of constant error variance against the alternative that the error variance changes with the level of the response (fitted values), or with a linear combination of predictors.
```

```
## Non-constant Variance Score Test
## Variance formula: ~ fitted.values
## Chisquare = 15.09463    Df = 1    p = 0.0001022536
```

```
# Conclusión: sigue existiendo hetercedasticidad, habrá que utilizar una estimación
# robusta de la matriz de varianzas y covarianzas
```



Proyecto de innovación  
“Innova Docencia” de la UCM  
Convocatoria 2018-2019  
Proyecto número 275

**R-adaptación de la asignatura de métodos econométricos en economía y finanzas del  
Grado de Estadística Aplicada**

Facultad de Estudios Estadísticos  
Departamento de Economía Aplicada, Pública y Política

**Título**

***Práctica 8 : Abracadabra. No estacionariedad en media y Cointegración***

**Correspondiente al**

**Capítulo 4 MODELOS CON DATOS DE SERIES TEMPORALES I: COINTEGRACIÓN Y  
REGRESIONES ESPURIAS**

- Magia “potagia” o las regresiones espurias
- Cointegración y modelos de regresión con **series no estacionarias**
- ¿Cuántas veces hay que diferenciar una serie para convertirla en estacionaria? El contraste de raíces unitarias
- Contrastes de cointegración
- Cointegración y Mecanismos de Corrección del Error (MCE)
  - Práctica 8 : Abracadabra: No estacionariedad e media y Cointegración***
  - Práctica 9 : Cointegración y Modelos de Corrección del Error***
  - Práctica 10: De vueltas con la ley de Okun: cointegración de las series de PIB y Empleo***

# Abracadabra, No Estacionariedad en media y Cointegración

*Asignatura*

*Métodos Econométricos en Economía y Finanzas*

*Profesores*

*Loren Escot, Alicia Pérez-Alonso y Julio E. Sandubete*

Dpto Economía Aplicada, Pública y Política

Facultad Estudios Estadísticos - Universidad Complutense de Madrid

*Curso académico 2018/2019*

## **Abstract**

En esta práctica se muestra un ejemplo donde la *no-estacionariedad* en media puede generar que aparezcan relaciones espúreas. Estas relaciones falsas aparecen cuando las series analizadas resultan no-estacionarias en media, por lo que se presentan los *TEST de RAICES UNITARIAS*, como test formales para contrastar la hipótesis alternativa de estacionariedad.

También se presentan tres alternativas para contrastar si las relaciones entre series no-estacionarias son espúreas o existe de hecho una relación de *COINTEGRACION* entre las series, esto es, una relación de largo plazo entre ellas.

- 1 Abracadabra: Vamos a hacer un truco de Magia
- 2 Cómo Detectar si una serie es estacionaria: test de Raíces Unitarias
- 3 Test de Cointegración

## 1 Abracadabra: Vamos a hacer un truco de Magia

```

# Vamos a hacer un truco de Magia

# A continuación vamos a generar dos series temporales seriex y seriey
# de forma totalmente independiente. Posteriormente utilizaré las palabras
# mágicas "abracadabra..." para generar correlación entre ellas

# genero la seriex (quiero que cada uno cambie los parámetros de la constante 1.6,
# y de la dt =0.5 para que cada uno tenga una serie distinta)

seriex0<-rnorm(1, mean=5, sd=1)
seriex<-rbind(seriex0)

for (i in 2:100){
  seriex_t<-1.6+seriex0+rnorm(1, mean=0, sd=0.5)
  seriex<-rbind(seriex, seriex_t)
  seriex0<-seriex_t}

# ahora genero la seriey (por favor cambiad la constante y dt para
# que cada uno tenga una serie distinta)

seriey0<-rnorm(1, mean=0, sd=2)
seriey<-rbind(seriey0)

for (i in 2:100){
  seriey_t<-1.5+seriey0+rnorm(1, mean=0, sd=4.5)
  seriey<-rbind(seriey, seriey_t)
  seriey0<-seriey_t}

#write.table(seriex,"clipboard", row.names = FALSE, col.names = FALSE)
#write.table(seriey,"clipboard", row.names = FALSE, col.names = FALSE)

# por construcción el amable público puede comprobar que son totalmente independientes
# además habeis cambiado los parámetros de la construcción de la series y
# por tanto cada uno tiene unas series distintas.....
# ..... atentos... concéntrense.... "abracadabra pata de cabra, que estas dos series
# independientes correlacionen perfectamente !!!! "

# ya está... comprueben:

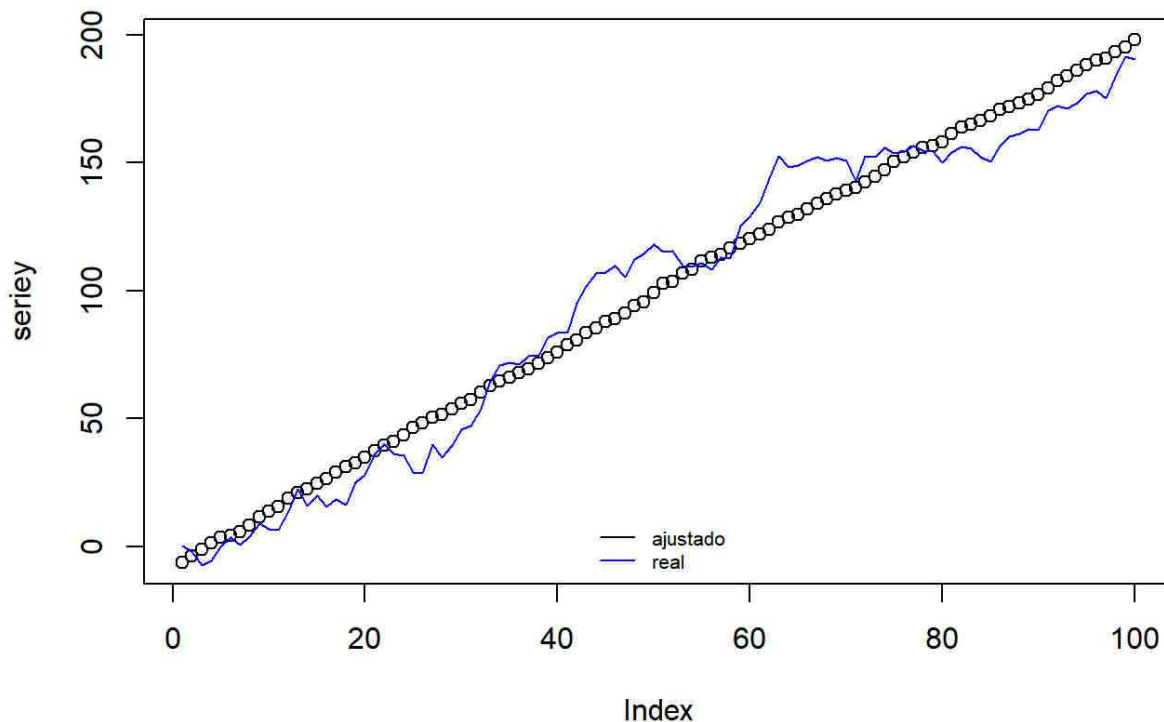
reg<-lm(seriey~seriex)
summary(reg)

```

```
##
## Call:
## lm(formula = seriey ~ seriex)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -19.889  -9.445  -1.628   7.782  25.916
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -13.71990     2.42660  -5.654 1.55e-07 ***
## seriex       1.35646     0.02596  52.260 < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 11.39 on 98 degrees of freedom
## Multiple R-squared:  0.9654, Adjusted R-squared:  0.965
## F-statistic: 2731 on 1 and 98 DF,  p-value: < 2.2e-16
```

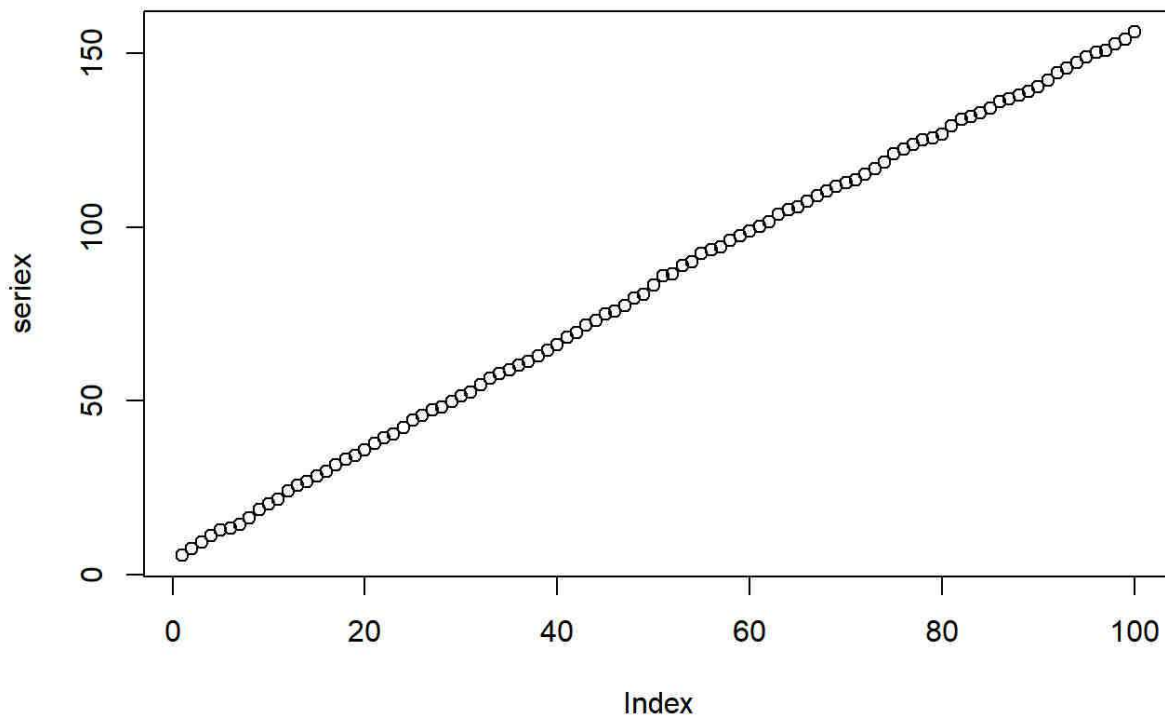
# como pueden ustedes comprobar el coeficiente de la regresión es  
# cercano a uno y altamente significativo, el R2 del 0.91 y el  
# estadístico F- de significación conjunta nos dice que el modelo es explicativo

```
plot(fitted(reg), ylab="seriey")
lines(seriey,col="blue")
legend("bottom",legend = c("ajustado", "real"),lty=c(1), cex=0.65, lwd=c(1,1),
      bty="n",
      col=c("black","blue"))
```



```
# ¿Qué es lo que ha pasado?  
# hemos correlacionado dos series que son NO ESTACIONARIAS en media y eso es  
# lo que ha hecho que aparezca esa correlación espúrea.
```

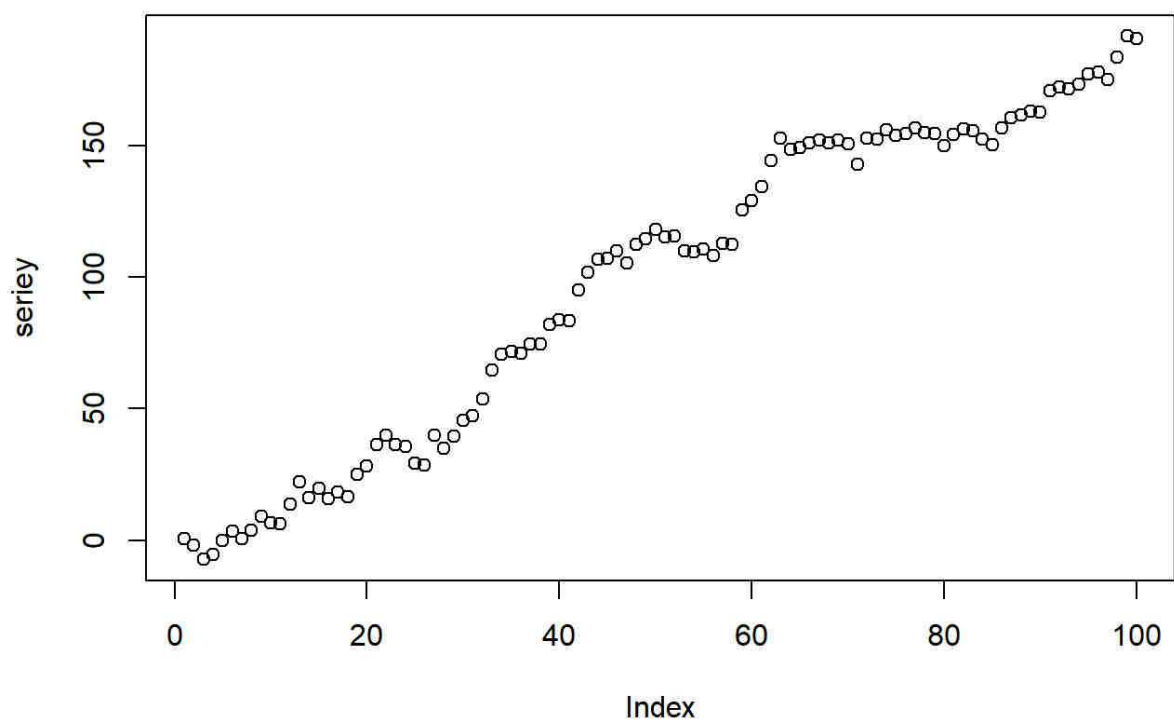
```
plot(serieX)
```



```
plot(serieY)
```

```
# Conclusión: Cuando se trabajan con series temporales hay que comprobar si  
# son espúreas o no, ya que si las series son NO-estacionarias en media  
# podría darse el caso de que la correlación (regresión significativa) sea  
# espúrea, de mentira, y que en realidad no haya una relación de verdad entre ellas  
# y todo sea debido a la falta de estacionariedad
```

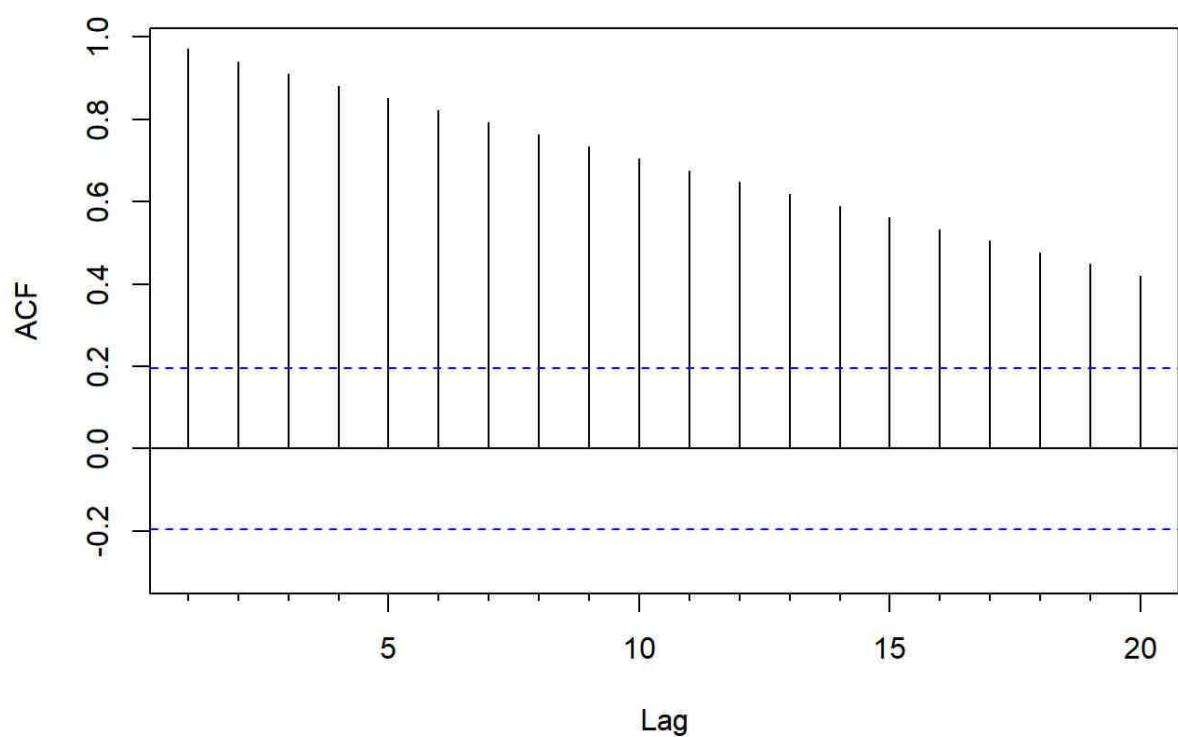
```
library(forecast)
```



# efectivamente el correlograma parece como el de una serie no estacionaria

```
Acf(seriey)
```

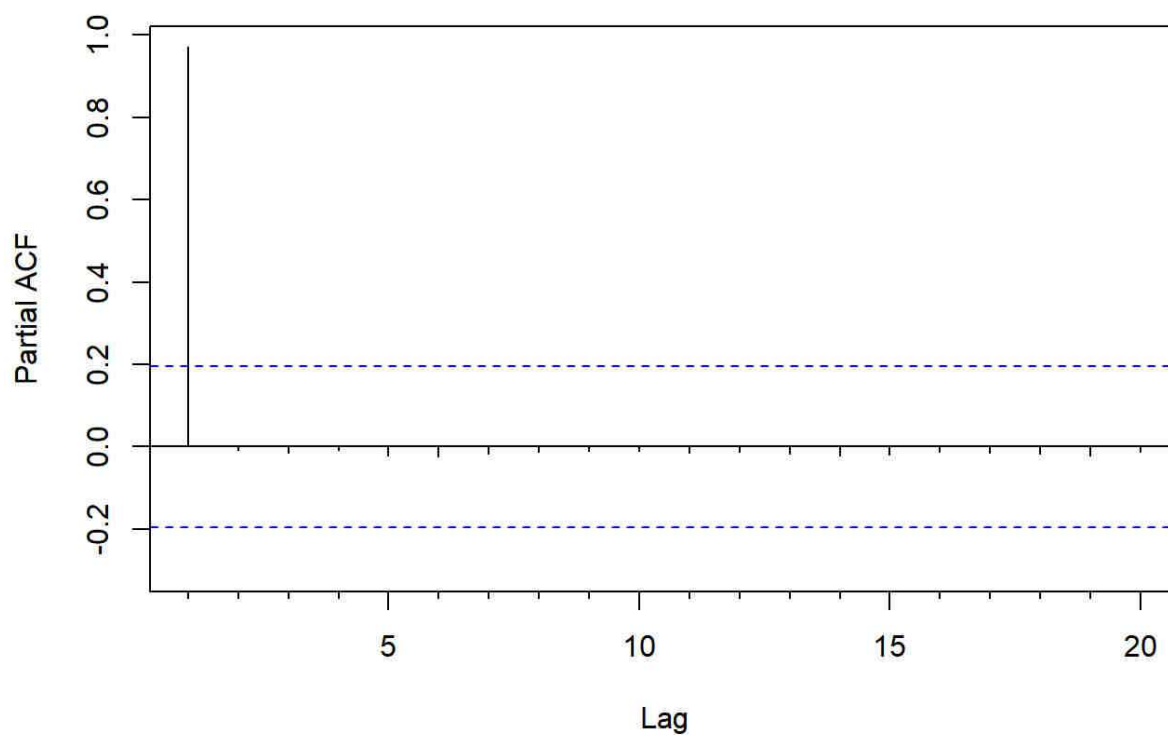
### Series seriey



```
Pacf(seriey)
```

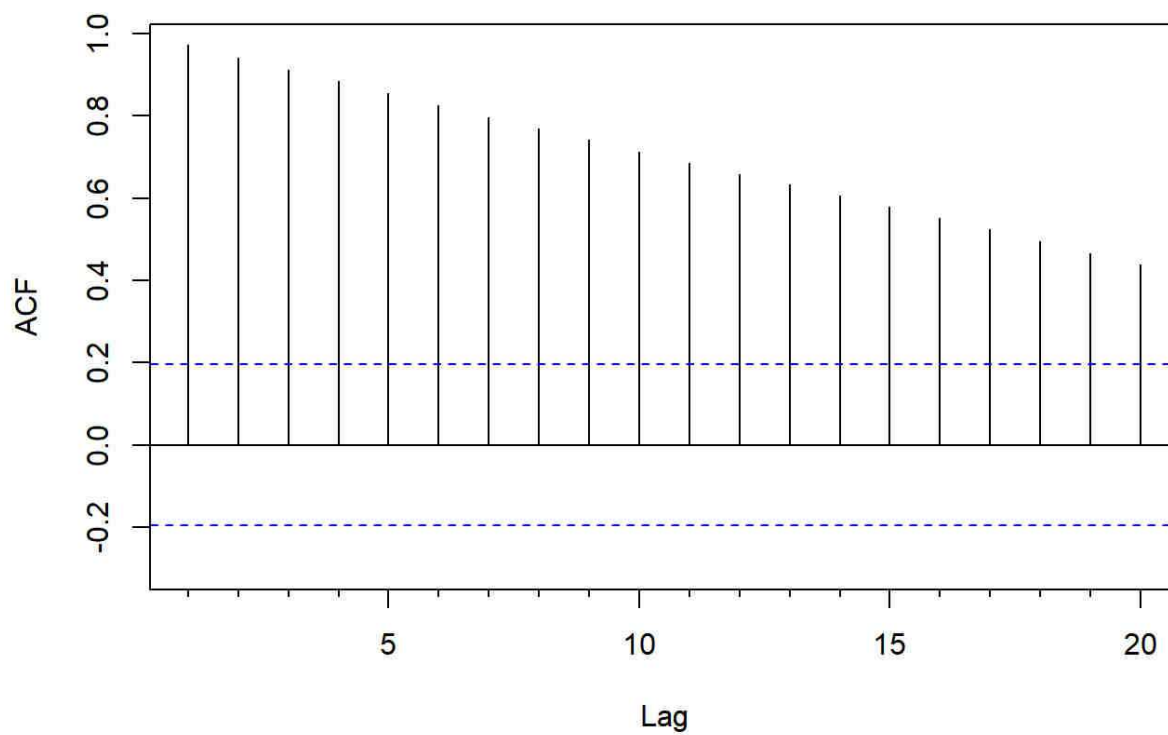


Series serieix



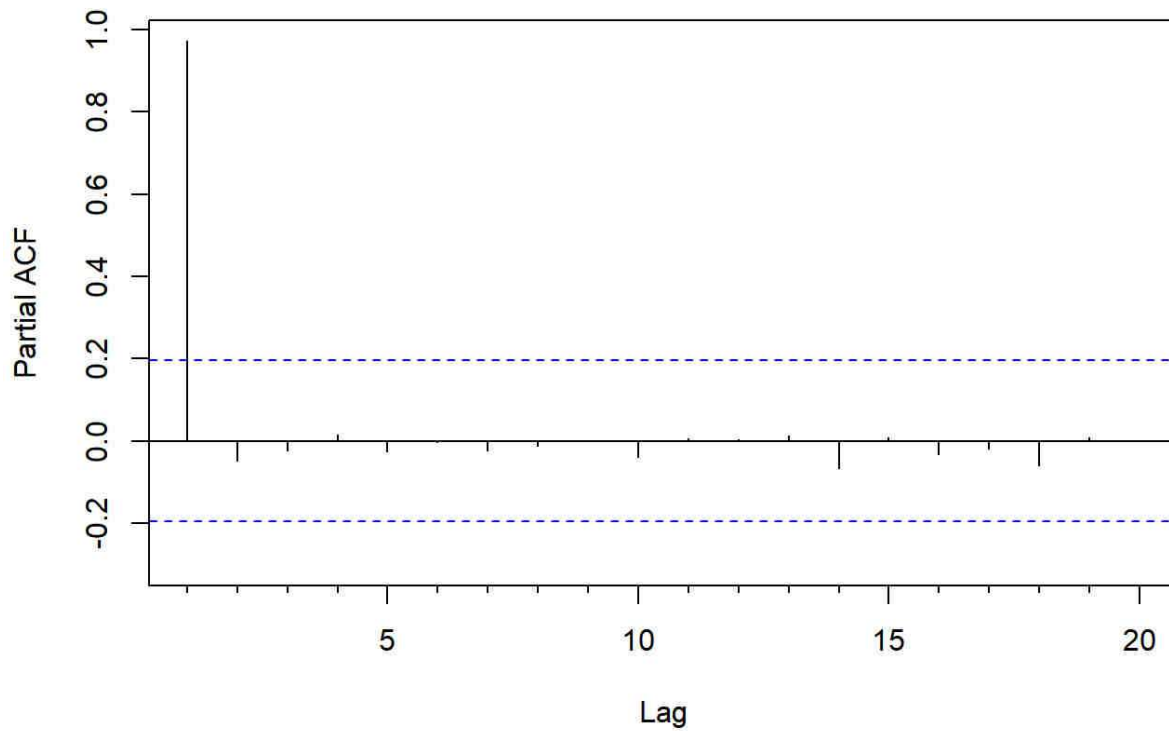
```
Acf(series)
```

Series seriey



```
Pacf(series)
```

## Series seriey



```
# otra pista de que a lo mejor la regresión es espurea es que el ajuste  
# de la regresión es muy muy bueno, pero los residuos no parecen estacionarios,  
# al contrario presentan también una autocorrelación muy muy fuerte
```

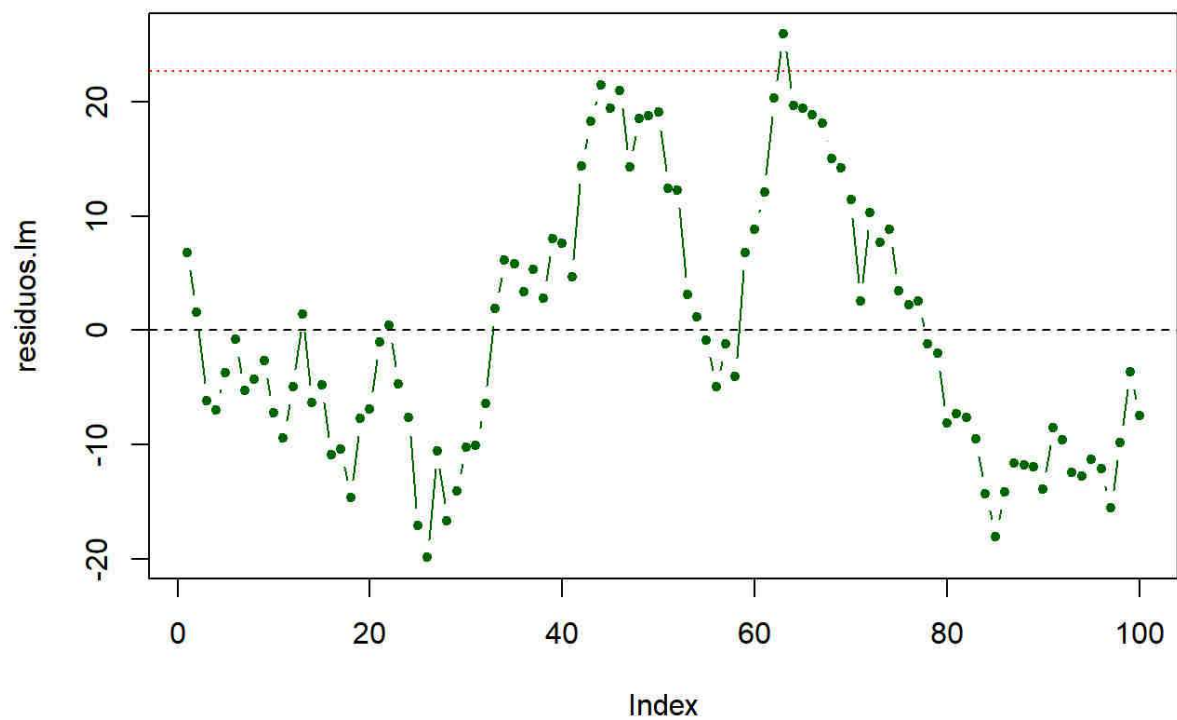
```
residuos.lm<-residuals(reg)
```

```
plot(residuos.lm, type="b", col="darkgreen",pch=20)
```

```
abline(h=0,lty=2, col="grey2" )
```

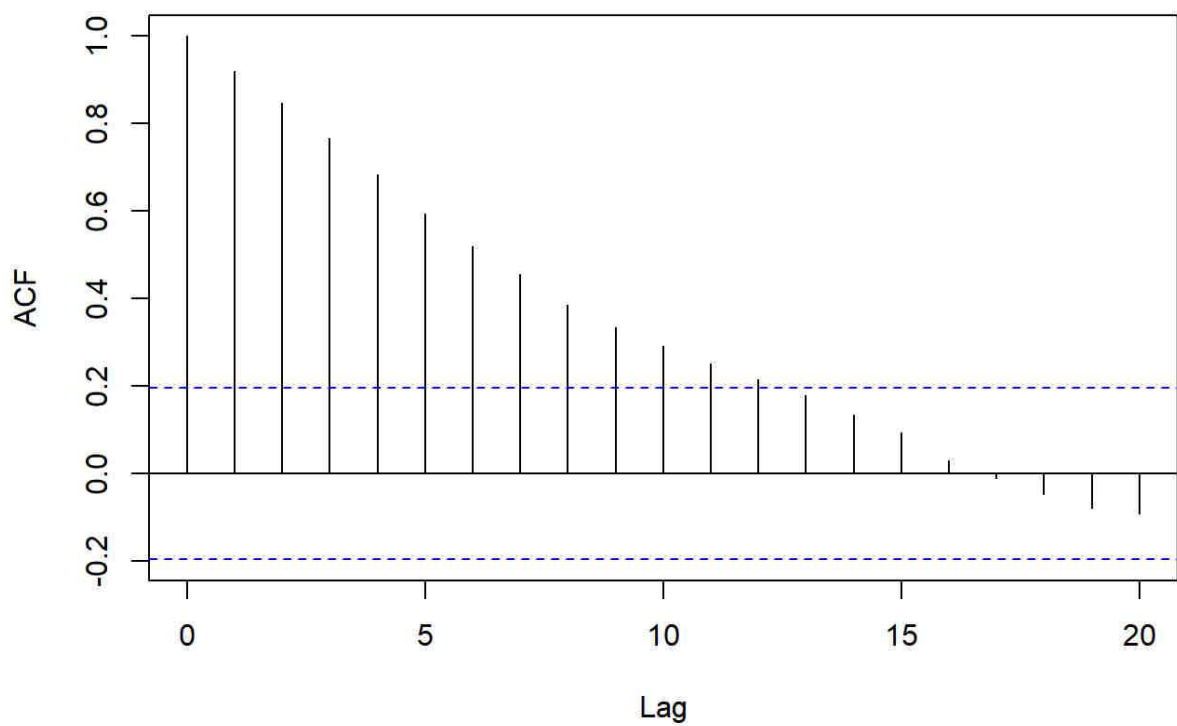
```
abline(h=2*sd(residuos.lm),lty=3, col="red" )
```

```
abline(h=-2*sd(residuos.lm),lty=3, col="red" )
```



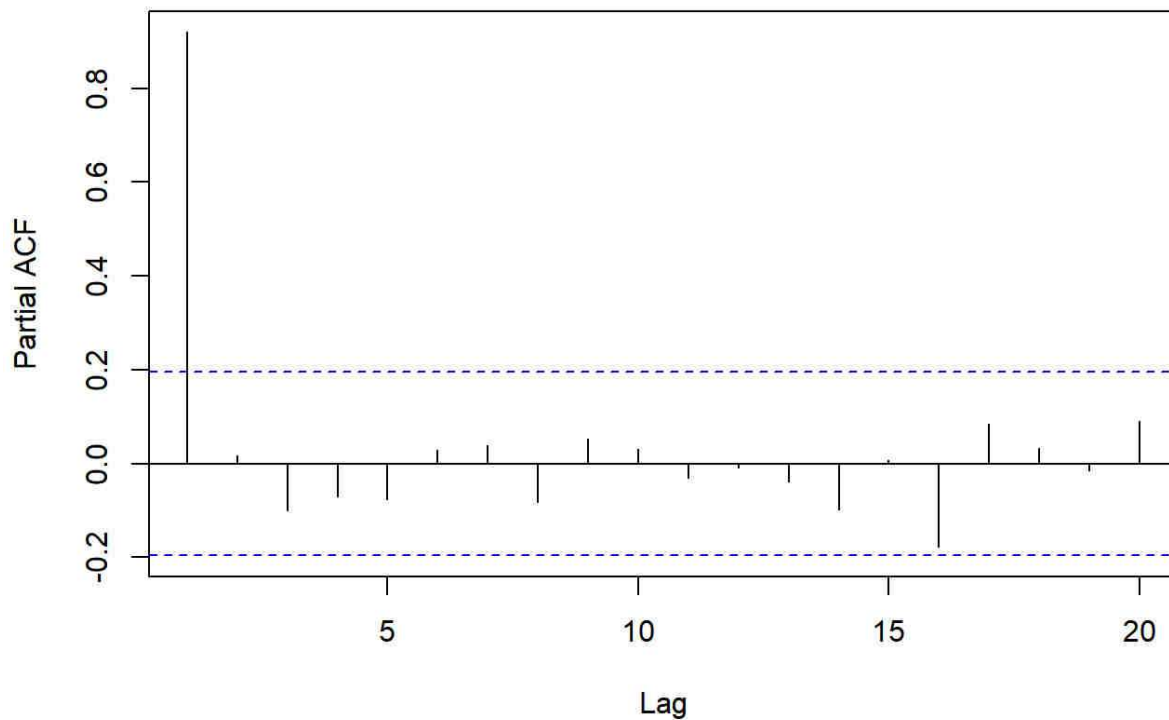
```
acf(residuos.lm)
```

### Series residuos.lm



```
pacf(residuos.lm)
```

## Series residuos.lm



```
library(lmtest)
```

```
## Warning: package 'lmtest' was built under R version 3.4.2
```

```
## Loading required package: zoo
```

```
##  
## Attaching package: 'zoo'
```

```
## The following objects are masked from 'package:base':  
##  
## as.Date, as.Date.numeric
```

```
dwtest(reg)
```

```
##  
## Durbin-Watson test  
##  
## data: reg  
## DW = 0.15242, p-value < 2.2e-16  
## alternative hypothesis: true autocorrelation is greater than 0
```

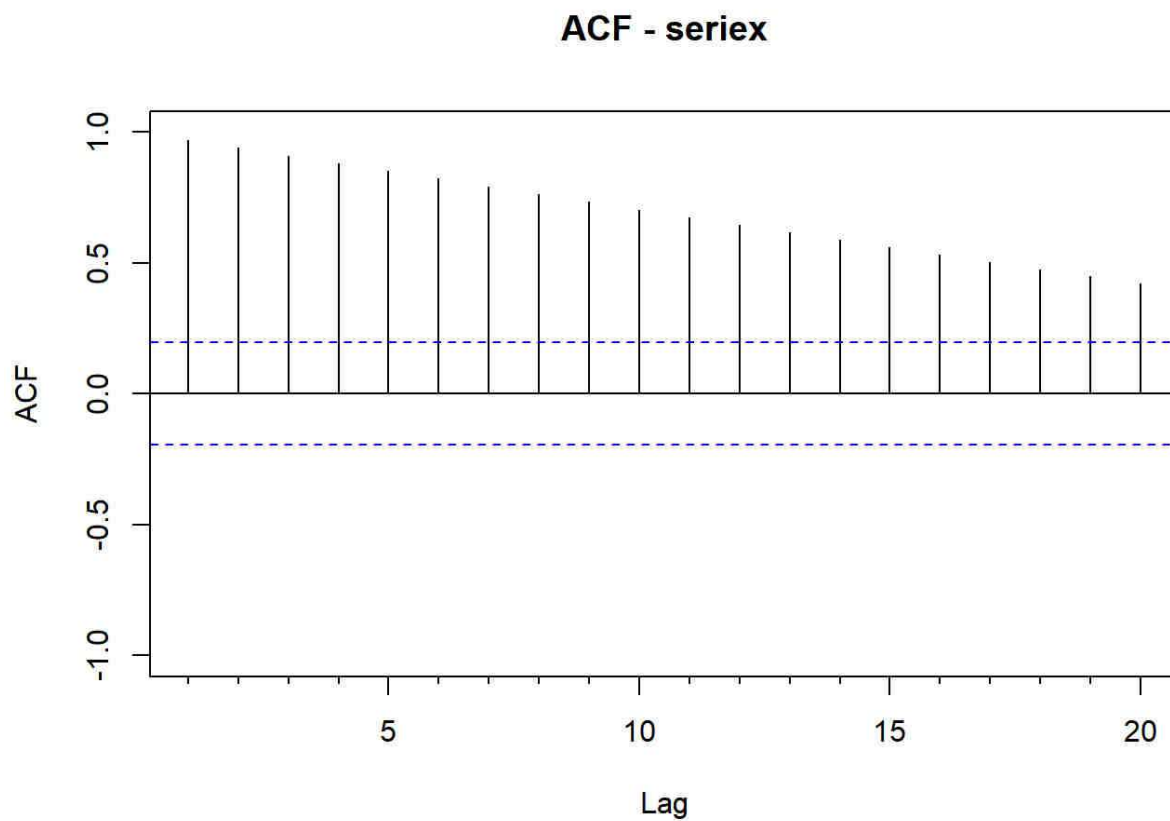
## 2 Cómo Detectar si una serie es estacionaria: test de Raíces Unitarias

```
# Cómo Detectar si una serie es estacionaria

#¿es la serie estacionaria?

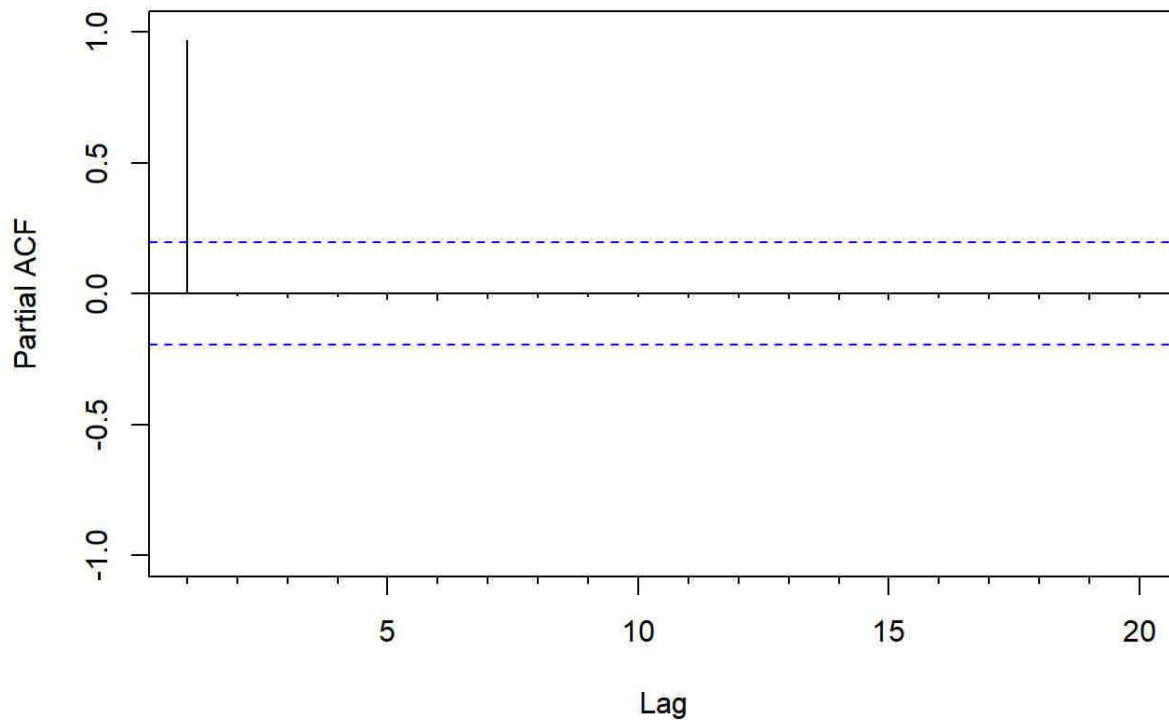
# primera opción Correlograma
library(forecast)

#par(mfrow=c(2,1))
Acf(x=serie,lag.max = 20, ylim=c(-1,1), xlim=c(1,20), main="ACF - serie")
```



```
Pacf(x=serie,lag.max = 20, ylim=c(-1,1),main="PACF - serie")
```

## PACF - seriex



```
#par(mfrow=c(1,1))
```

```
# Existe sin embargo un test formal: test de raices unitarias
library(tseries)
adf.test(na.omit(seriex))
```

```
##
## Augmented Dickey-Fuller Test
##
## data: na.omit(seriex)
## Dickey-Fuller = -0.85876, Lag order = 4, p-value = 0.9536
## alternative hypothesis: stationary
```

```
adf.test(na.omit(seriey))
```

```
##
## Augmented Dickey-Fuller Test
##
## data: na.omit(seriey)
## Dickey-Fuller = -1.8918, Lag order = 4, p-value = 0.6214
## alternative hypothesis: stationary
```

```
ndiffs(na.omit(seriex)) #Number of differences required to achieve stationarity (from the forecast p
ackage)
```

```
## [1] 1
```

```
ndiffs(na.omit(seriey)) #Number of differences required to achieve stationarity (from the forecast package)
```

```
## [1] 1
```

```
adf.test(seriey)# Ho: una raíz unitaria (no estacionaria)
```

```
##  
## Augmented Dickey-Fuller Test  
##  
## data: seriey  
## Dickey-Fuller = -0.85876, Lag order = 4, p-value = 0.9536  
## alternative hypothesis: stationary
```

```
pp.test(seriey)
```

```
##  
## Phillips-Perron Unit Root Test  
##  
## data: seriey  
## Dickey-Fuller Z(alpha) = -7.1763, Truncation lag parameter = 3,  
## p-value = 0.7  
## alternative hypothesis: stationary
```

```
ndiffs(seriey) # directamente nos informa sobre el número de diferencias necesaria para convertir la serie
```

```
## [1] 1
```

```
# otra librería: librería URCA  
library(urca)#
```

```
## Warning: package 'urca' was built under R version 3.4.4
```

```
# augmented dickey-Fuller test ur.df , type = c("none", "drift", "trend")  
# Ho: una raíz unitaria (no estacionaria)  
tru<-ur.df(seriey, type = c("trend"), lags = 12,  
          selectlags = c("BIC"))  
summary(tru)
```

```
##
## #####
## # Augmented Dickey-Fuller Test Unit Root Test #
## #####
##
## Test regression trend
##
##
## Call:
## lm(formula = z.diff ~ z.lag.1 + 1 + tt + z.diff.lag)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -9.3793 -2.8349 -0.4835  2.7564 10.7217
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  2.29027     1.22499   1.870   0.0651 .
## z.lag.1      -0.05894     0.03863  -1.526   0.1309
## tt           0.11028     0.08012   1.377   0.1724
## z.diff.lag   -0.01239     0.10887  -0.114   0.9097
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 4.523 on 83 degrees of freedom
## Multiple R-squared:  0.03086,    Adjusted R-squared:  -0.004166
## F-statistic: 0.8811 on 3 and 83 DF,  p-value: 0.4544
##
##
## Value of test-statistic is: -1.5257 5.6064 1.2568
##
## Critical values for test statistics:
##      1pct  5pct 10pct
## tau3 -4.04 -3.45 -3.15
## phi2  6.50  4.88  4.16
## phi3  8.73  6.49  5.47
```

# si el valor del estadístico es mayor (ojo, respetando el signo) que en valor en tablas NO se rechaza la hipótesis nula,  
# por el contrario, si el estadístico es menor (no en valor absoluto sino más negativo), sí se rechaza  
# la hipótesis nula (raíz unitaria, no estacionariedad).

```
punitroot(tru@teststat[1], trend="ct")
```

```
## [1] 0.8210797
```

```
tru<-ur.df(seriey, type = c("drift"), lags = 12,
           selectlags = c("BIC"))
summary(tru)
```



```
##
## #####
## # Augmented Dickey-Fuller Test Unit Root Test #
## #####
##
## Test regression drift
##
##
## Call:
## lm(formula = z.diff ~ z.lag.1 + 1 + z.diff.lag)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -9.5951 -2.8272 -0.2963  2.5217 10.7668
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  2.824434   1.168080   2.418   0.0178 *
## z.lag.1      -0.007337   0.009376  -0.782   0.4361
## z.diff.lag   -0.039780   0.107606  -0.370   0.7125
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 4.547 on 84 degrees of freedom
## Multiple R-squared:  0.008738,    Adjusted R-squared:  -0.01486
## F-statistic: 0.3702 on 2 and 84 DF,  p-value: 0.6917
##
##
## Value of test-statistic is: -0.7825 7.3835
##
## Critical values for test statistics:
##      1pct  5pct 10pct
## tau2 -3.51 -2.89 -2.58
## phi1  6.70  4.71  3.86
```

```
punitroot(tru@teststat[1], trend="c")
```

```
## [1] 0.8234993
```

```
tru<-ur.df(seriey, type = c("none"), lags = 12,
           selectlags = c("BIC"))
summary(tru)
```

```
##
## #####
## # Augmented Dickey-Fuller Test Unit Root Test #
## #####
##
## Test regression none
##
##
## Call:
## lm(formula = z.diff ~ z.lag.1 - 1 + z.diff.lag)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -9.7518 -2.4947 -0.2793  3.4815 11.3373
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## z.lag.1      0.012820   0.004413   2.905  0.00468 **
## z.diff.lag  0.012652   0.108361   0.117  0.90733
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 4.674 on 85 degrees of freedom
## Multiple R-squared:  0.1057, Adjusted R-squared:  0.08467
## F-statistic: 5.024 on 2 and 85 DF, p-value: 0.008664
##
##
## Value of test-statistic is: 2.905
##
## Critical values for test statistics:
##      1pct  5pct 10pct
## tau1 -2.6 -1.95 -1.61
```

```
punitroot(tru@teststat[1], trend="nc")
```

```
## [1] 0.9992301
```

```
# otra librería
library(fUnitRoots)
```

```
## Warning: package 'fUnitRoots' was built under R version 3.4.4
```

```
## Loading required package: timeDate
```

```
## Loading required package: timeSeries
```

```
##
## Attaching package: 'timeSeries'
```

```
## The following object is masked from 'package:zoo':
##
##      time<-
```

```
## Loading required package: fBasics
```

```
##
```

```
## Rmetrics Package fBasics
```

```
## Analysing Markets and calculating Basic Statistics
```

```
## Copyright (C) 2005-2014 Rmetrics Association Zurich
```

```
## Educational Software for Financial Engineering and Computational Science
```

```
## Rmetrics is free software and comes with ABSOLUTELY NO WARRANTY.
```

```
## https://www.rmetrics.org --- Mail to: info@rmetrics.org
```

```
##
```

```
## Attaching package: 'fUnitRoots'
```

```
## The following objects are masked from 'package:urca':
```

```
##
```

```
##      punitroot, qunitroot, unitrootTable
```

```
adfTest(seriey, lags=4, type=c("ct"))
```

```
##
```

```
## Title:
```

```
## Augmented Dickey-Fuller Test
```

```
##
```

```
## Test Results:
```

```
##   PARAMETER:
```

```
##     Lag Order: 4
```

```
##   STATISTIC:
```

```
##     Dickey-Fuller: -1.8918
```

```
##   P VALUE:
```

```
##     0.6214
```

```
##
```

```
## Description:
```

```
## Sun Jun 30 09:21:29 2019 by user: loren
```

```
unitrootTest(seriey, lags = 4, type=c("ct"))
```

```
##
## Title:
## Augmented Dickey-Fuller Test
##
## Test Results:
## PARAMETER:
## Lag Order: 4
## STATISTIC:
## DF: -1.8918
## P VALUE:
## t: 0.6513
## n: 0.9714
##
## Description:
## Sun Jun 30 09:21:29 2019 by user: loren
```

```
# para conocer cuantas veces necesito diferenciar la serie para
# convertirla en estacionaria debo diferenciar la serie y pasar
# el test de raices unitarias hasta que éste me dice que la serie
# no tiene más raices unitarias y que por tanto ya es estacionaria

# diferencio una vez y consigo estacionariedad
tru<-ur.df(diff(serieX, lag=1), type = c("trend"), lags = 12,
          selectlags = c("BIC"))
summary(tru)
```

```
##
## #####
## # Augmented Dickey-Fuller Test Unit Root Test #
## #####
##
## Test regression trend
##
##
## Call:
## lm(formula = z.diff ~ z.lag.1 + 1 + tt + z.diff.lag)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.9819 -0.3079 -0.0171  0.2688  1.3786
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  2.052039   0.300365   6.832 1.36e-09 ***
## z.lag.1      -1.270545   0.168233  -7.552 5.35e-11 ***
## tt           -0.002567   0.002149  -1.195   0.236
## z.diff.lag    0.093014   0.110654   0.841   0.403
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.488 on 82 degrees of freedom
## Multiple R-squared:  0.582, Adjusted R-squared:  0.5667
## F-statistic: 38.05 on 3 and 82 DF, p-value: 1.655e-15
##
##
## Value of test-statistic is: -7.5523 19.022 28.5197
##
## Critical values for test statistics:
##      1pct  5pct 10pct
## tau3 -4.04 -3.45 -3.15
## phi2  6.50  4.88  4.16
## phi3  8.73  6.49  5.47
```

```
punitroot(tru@teststat[1], trend="ct")
```

```
## [1] 5.391686e-11
```

```
tru<-ur.df(diff(seriex, lag=1), type = c("drift"), lags = 12,
           selectlags = c("BIC"))
summary(tru)
```

```
##
## #####
## # Augmented Dickey-Fuller Test Unit Root Test #
## #####
##
## Test regression drift
##
##
## Call:
## lm(formula = z.diff ~ z.lag.1 + 1 + z.diff.lag)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.08314 -0.27781 -0.01887  0.21738  1.38069
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  1.86015     0.25445   7.310 1.51e-10 ***
## z.lag.1      -1.23752     0.16637  -7.438 8.45e-11 ***
## z.diff.lag    0.07693     0.11011   0.699  0.487
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.4893 on 83 degrees of freedom
## Multiple R-squared:  0.5747, Adjusted R-squared:  0.5644
## F-statistic: 56.07 on 2 and 83 DF, p-value: 3.904e-16
##
##
## Value of test-statistic is: -7.4383 27.6772
##
## Critical values for test statistics:
##      1pct  5pct 10pct
## tau2 -3.51 -2.89 -2.58
## phi1  6.70  4.71  3.86
```

```
punitroot(tru@teststat[1], trend="c")
```

```
## [1] 2.051109e-11
```

### 3 Test de Cointegración

```
# ¿Cómo detectar si una regresión con series NO-Estacionarias es espurea?

# Opción 1 trasformo las series para que sean estacionarias y vuelvo
# a hacer la regresión con las series estacionarias. Si la relación inicial era de
# verdad ésta nueva regresión con series estacionarias seguirá siendo
# significativa. Si por el contrario, desaparece la relación entre ellas cuando
# las hemos convertido en estacionarias, es que la regresión inicial era espurea.

# construyo las series diferenciadas dy dx
dy <- diff(seriey)
dx <- diff(seriex)

# nueva regresión con las series estacionarias
dif.reg<-lm(dy~dx)
summary(dif.reg)
```



```
# Opción 2: Test cointegración Engle-Granger

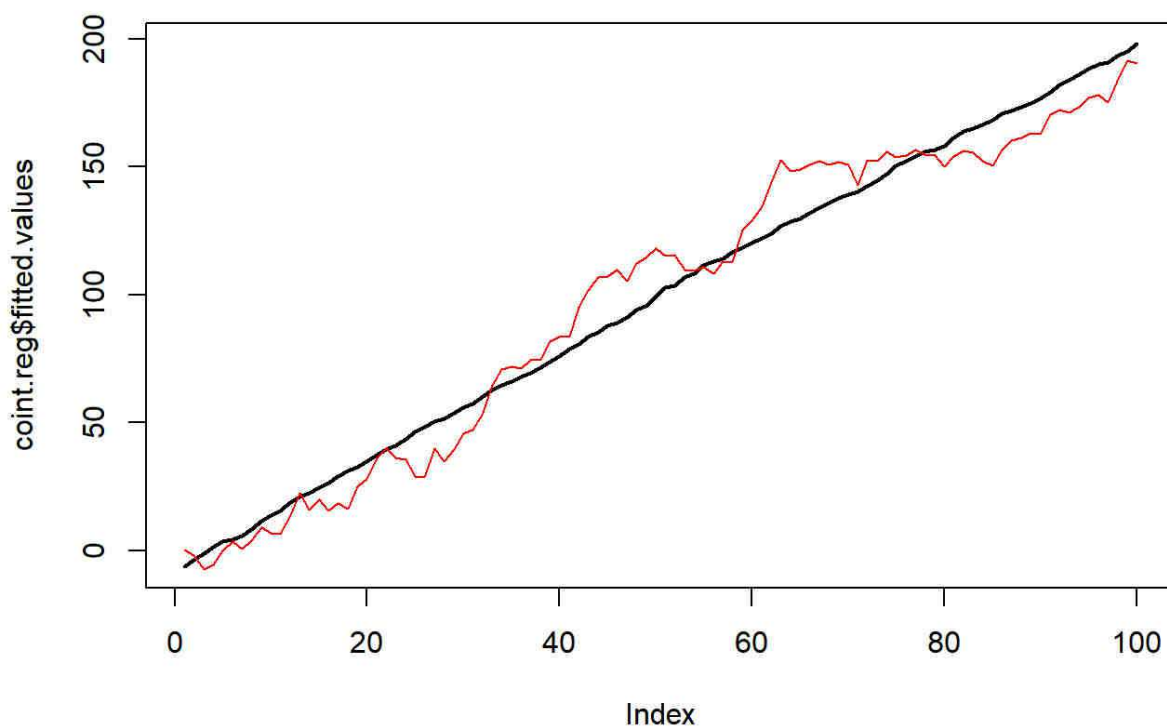
# Este es un test formal de cointegración, de relación de verdad a largo plazo.

# Primera etapa hago la relación entre las dos series NO-Estacionarias

# coint.reg <- lm(series ~ serieX -1 )
coint.reg <- lm(series ~ serieX )
summary(coint.reg)
```

```
##
## Call:
## lm(formula = series ~ serieX)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -19.889  -9.445  -1.628   7.782  25.916
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -13.71990    2.42660  -5.654 1.55e-07 ***
## serieX       1.35646    0.02596  52.260 < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 11.39 on 98 degrees of freedom
## Multiple R-squared:  0.9654, Adjusted R-squared:  0.965
## F-statistic: 2731 on 1 and 98 DF, p-value: < 2.2e-16
```

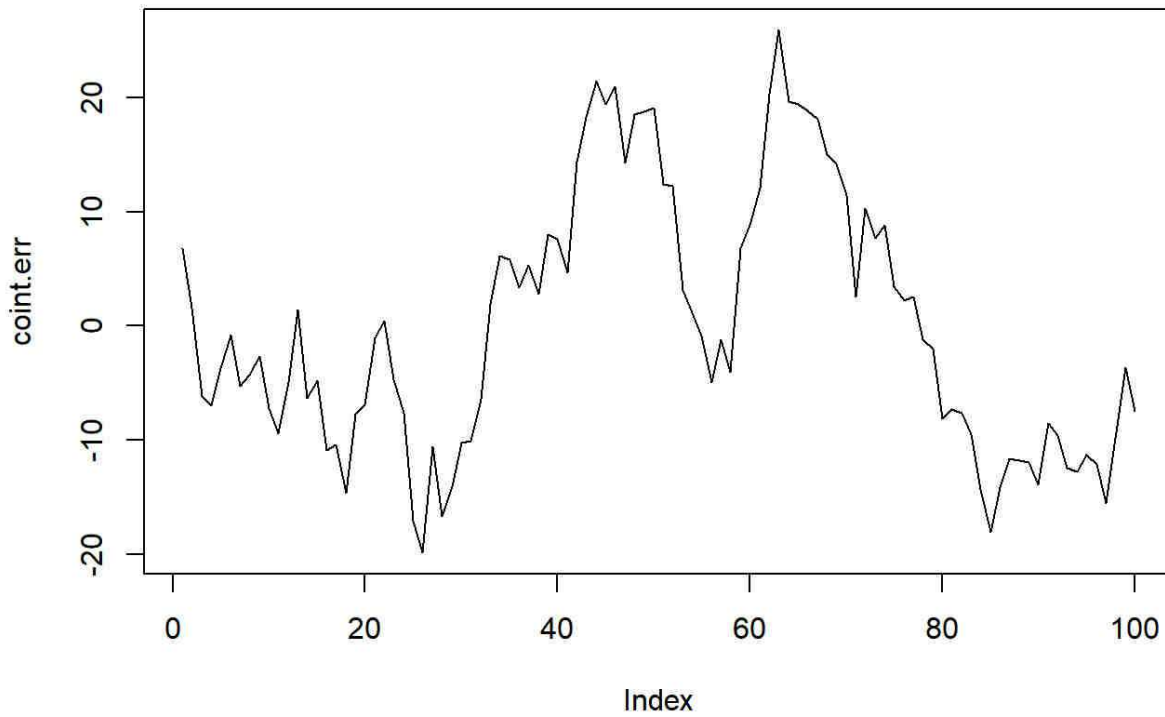
```
plot(coint.reg$fitted.values, type="l", lwd=2)
lines(series, type="l", col=2)
```





```
# extraigo el error
coint.err<-residuals(coint.reg)

plot(coint.err, type="l")
```



```
# segunda etapa: compruebo si el erroer es estacionario
library(fUnitRoots)

adf.test(coint.err) # ojo el p-valor no es v?lido por eso tengo que calcular el p-valor de manera ro
busta:
```

```
##
## Augmented Dickey-Fuller Test
##
## data: coint.err
## Dickey-Fuller = -2.1266, Lag order = 4, p-value = 0.5242
## alternative hypothesis: stationary
```

```
urca::punitroot(adf.test(coint.err)$statistic, trend="ct")
```

```
## Dickey-Fuller
## 0.530317
```

```
tru<-ur.df(coint.err, type = c("trend"), lags = 12,
          selectlags = c("BIC"))
summary(tru)
```

```
##
## #####
## # Augmented Dickey-Fuller Test Unit Root Test #
## #####
##
## Test regression trend
##
##
## Call:
## lm(formula = z.diff ~ z.lag.1 + 1 + tt + z.diff.lag)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -10.260  -2.911  -0.489   2.826  10.545
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  0.41620    1.17718   0.354   0.725
## z.lag.1      -0.06685    0.04122  -1.622   0.109
## tt           -0.00856    0.01916  -0.447   0.656
## z.diff.lag   -0.03137    0.10892  -0.288   0.774
##
## Residual standard error: 4.471 on 83 degrees of freedom
## Multiple R-squared:  0.03572,    Adjusted R-squared:  0.0008675
## F-statistic: 1.025 on 3 and 83 DF,  p-value: 0.3859
##
##
## Value of test-statistic is: -1.6216 0.9262 1.3668
##
## Critical values for test statistics:
##      1pct  5pct 10pct
## tau3 -4.04 -3.45 -3.15
## phi2  6.50  4.88  4.16
## phi3  8.73  6.49  5.47
```

```
tru@teststat[1]
```

```
## [1] -1.621577
```

```
punitroot(tru@teststat[1], trend="ct")
```

```
## [1] 0.7848194
```

```
# si el valor del estadístico es mayor (ojo, respetando el signo) que en valor en tablas NO se rechaza la hipótesis nula,  
# por el contrario, si el estadístico es menor (no en valor absoluto sino más negativo), sí se rechaza  
# la hipótesis nula (raíz unitaria en los residuos, no estacionariedad de los residuos, no cointegración de las series).
```

```
# Phillips-Ouliaris Cointegration Test H0: NO Cointegration
```

```
# null hypothesis that x is not cointegrated.
```

```
series<-matrix(data=c(seriey,seriex),nrow=length(seriey),ncol=2)
```

```
seriey.ts<-ts(seriey, start=1, frequency=1)
```

```
seriex.ts<-ts(seriex, start=1, frequency=1)
```

```
series<-ts.intersect(seriey.ts,seriex.ts)
```

```
library(tseries)
```

```
po.test(series,demean = FALSE, lshort=TRUE) #Phillips-Ouliaris Cointegration Test H0: NO Cointegration
```

```
## Warning in po.test(series, demean = FALSE, lshort = TRUE): p-value greater  
## than printed p-value
```

```
##
```

```
## Phillips-Ouliaris Cointegration Test
```

```
##
```

```
## data: series
```

```
## Phillips-Ouliaris standard = -5.7935, Truncation lag parameter =
```

```
## 0, p-value = 0.15
```

```
library(urca)
```

```
# null hypothesis that x is not cointegrated.
```

```
capocointest<-ca.po(series,demean=c("constant"),lag=c("short"))
```

```
summary(capocointest)
```

```
##
## #####
## # Phillips and Ouliaris Unit Root Test #
## #####
##
## Test of type Pu
## detrending of series with constant only
##
##
## Call:
## lm(formula = z[, 1] ~ z[, -1])
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -19.889  -9.445  -1.628   7.782  25.916
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -13.71990     2.42660  -5.654 1.55e-07 ***
## z[, -1]      1.35646     0.02596  52.260 < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 11.39 on 98 degrees of freedom
## Multiple R-squared:  0.9654, Adjusted R-squared:  0.965
## F-statistic: 2731 on 1 and 98 DF, p-value: < 2.2e-16
##
##
## Value of test-statistic is: 16.3302
##
## Critical values of Pu are:
##              10pct   5pct   1pct
## critical values 27.8536 33.713 48.0021
```

```
# si el estadístico es mayor que en valor en tabla rechazamos la Ho nula
```

```
# si el valor del estadístico es mayor (ojo, respetando el signo) que en valor en tablas NO se rechaza la hipótesis nula,
# por el contrario, si el estadístico es menor (no en valor absoluto sino más negativo), sí se rechaza
# la hipótesis nula (raíz unitaria, no estacionariedad, no cointegración).
```

```
# Opción 3: Test de Johansen
# In Johansen cointegration test, the null hypothesis for
# the eigenvalue test is that there are cointegration relations.
# The test is therefore sequential: you test first for r=0, then r=1, etc.
# The test concludes on the value of when the test fails
# to reject for the first time.
```

```
coint.test1 <- ca.jo(series,
                      ecdet = "trend",
                      type="trace",
                      K=2,
                      spec="longrun"
)
summary(coint.test1) # Check the cointegration rank here
```

```
##
## #####
## # Johansen-Procedure #
## #####
##
## Test type: trace statistic , with linear trend in cointegration
##
## Eigenvalues (lambda):
## [1] 9.996308e-02 2.852051e-02 -2.082812e-17
##
## Values of teststatistic and critical values of test:
##
##          test 10pct 5pct 1pct
## r <= 1 | 2.84 10.49 12.25 16.26
## r = 0 | 13.16 22.76 25.32 30.45
##
## Eigenvectors, normalised to first column:
## (These are the cointegration relations)
##
##          seriey.ts.l2 seriex.ts.l2 trend.l2
## seriey.ts.l2      1.00000      1.000000      1.00000
## seriex.ts.l2     -10.73181      4.062662 -55.10725
## trend.l2         14.11863     -8.010905  85.08678
##
## Weights W:
## (This is the loading matrix)
##
##          seriey.ts.l2 seriex.ts.l2      trend.l2
## seriey.ts.d -0.03674706 -0.04258488 -1.334628e-15
## seriex.ts.d  0.01270986 -0.00210801  5.377169e-16
```

```
coint.test1@cval
```

```
##          10pct 5pct 1pct
## r <= 1 | 10.49 12.25 16.26
## r = 0 | 22.76 25.32 30.45
```

```
coint.test1@teststat
```

```
## [1] 2.835642 13.156953
```

```
# si el valor del estadístico es mayor (positivo) que en valor en tablas se rechaza la hipótesis nula
# de que existe r relaciones de cointegración. Por el contrario, si el estadístico
# es menor que los valores tabulados entonces sí se rechaza
# la hipótesis nula (de r relaciones de cointegración).
# r es el número de relaciones de cointegración
```



Proyecto de innovación  
“Innova Docencia” de la UCM  
Convocatoria 2018-2019  
Proyecto número 275

**R-adaptación de la asignatura de métodos econométricos en economía y finanzas del  
Grado de Estadística Aplicada**

Facultad de Estudios Estadísticos  
Departamento de Economía Aplicada, Pública y Política

**Título**

***Práctica 9 : Cointegración y Modelos de Corrección del Error***

**Correspondiente al**

**Capítulo 4 MODELOS CON DATOS DE SERIES TEMPORALES I: COINTEGRACIÓN Y  
REGRESIONES ESPURIAS**

- Magia “potagia” o las regresiones espurias
- Cointegración y modelos de regresión con **series no estacionarias**
- ¿Cuántas veces hay que diferenciar una serie para convertirla en estacionaria? El contraste de raíces unitarias
- Contrastes de cointegración
- Cointegración y Mecanismos de Corrección del Error (MCE)
  - Práctica 8 : Abracadabra: No estacionariedad e media y Cointegración***
  - Práctica 9 : Cointegración y Modelos de Corrección del Error***
  - Práctica 10: De vueltas con la ley de Okun: cointegración de las series de PIB y Empleo***

# Cointegración y Modelos de Corrección del Error

## *Asignatura* *Métodos Econométricos en Economía y Finanzas*

*Profesores*  
*Loren Escot, Alicia Pérez-Alonso y Julio E. Sandubete*

Dpto Economía Aplicada, Pública y Política  
Facultad Estudios Estadísticos - Universidad Complutense de Madrid

*Curso académico 2018/2019*

### **Abstract**

En esta práctica seguimos con el análisis de *cointegración* de series no estacionarias en media.

Seguiremos un ejemplo en el que dos series no-estacionarias sí están cointegradas.

Además se presentarán los modelos de corrección del error. Estos modelos sirven para mejorar el ajuste a corto plazo entre series no-estacionarias que se sabe que están cointegradas, y permiten por tanto mejorar las predicciones.

- 1 Generamos dos series no-estacionarias en media que **sí** están cointegradas
- 2 ANALISIS DE COINTEGRACION
  - 2.1 Estacionariedad
    - 2.1.1 test de Raíces Unitarias
- 3 Test de Cointegración
  - 3.1 Opción 1: Regresión sobre series diferenciadas para conseguir estacionariedad
  - 3.2 Opción 2: Test cointegración Engle-Granger
    - 3.2.1 Primera etapa
    - 3.2.2 extraigo el error
    - 3.2.3 segunda etapa: compruebo si el error es estacionario
    - 3.2.4 Phillips–Ouliaris Cointegration Test  $H_0$ : NO Cointegration
  - 3.3 Opción 3: Test de Johansen
- 4 Modelo de Corrección del Error

```
# Test de Engle Granger de cointegración y  
# estimación del modelo de Corrección de Error
```

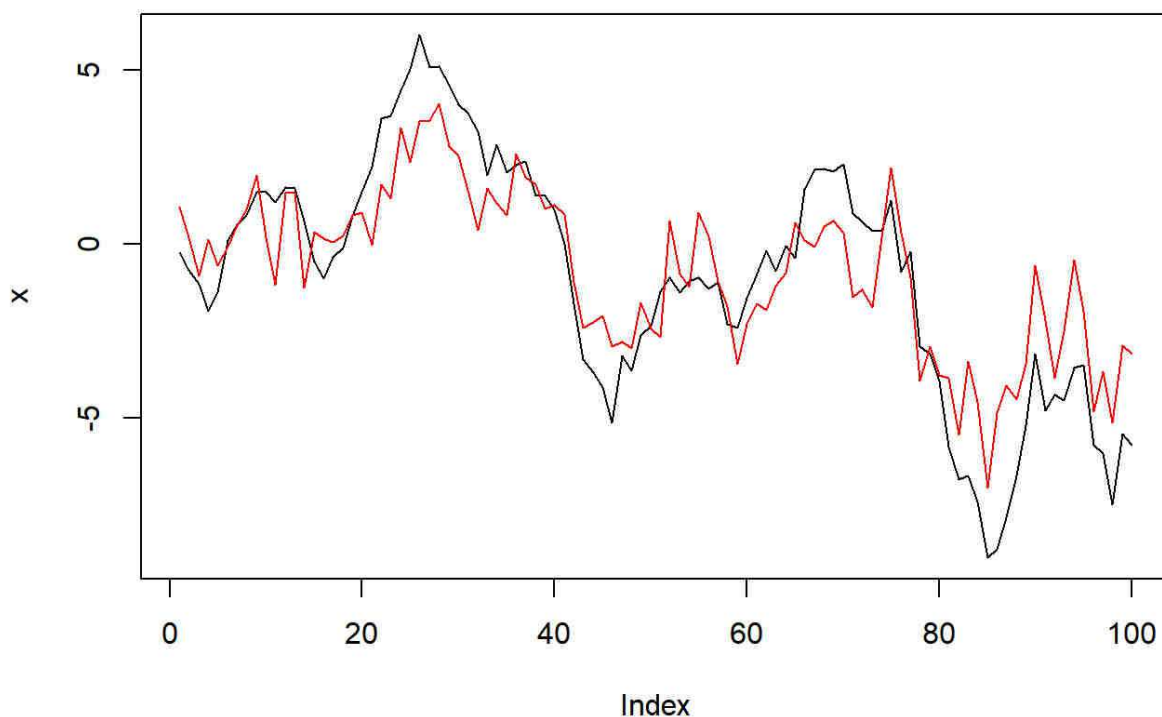
## 1 Generamos dos series no-estacionarias en media que **sí** están cointegradas

```
# creo dos series cointegradas (RELACION NO ESPUREA)
# xt= xt-1 + et
# yt= 0.6*xt + vt

set.seed(56)

e <- rnorm(100)
v <- rnorm(100)
x <- cumsum(+e)
y <- 0.6*x + v

plot(x, type="l", ylim=c(min(x,y),max(x,y)))
lines(y, type="l", col=2)
```



```
#write.table(x,"clipboard", row.names = FALSE, col.names = FALSE)
#write.table(y,"clipboard", row.names = FALSE, col.names = FALSE)
```

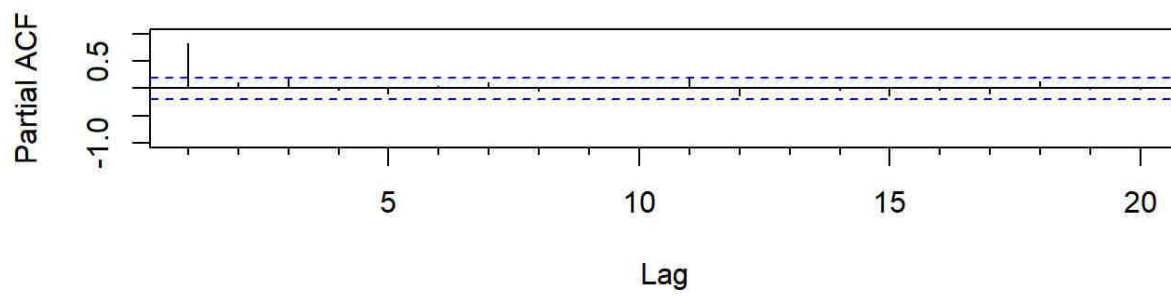
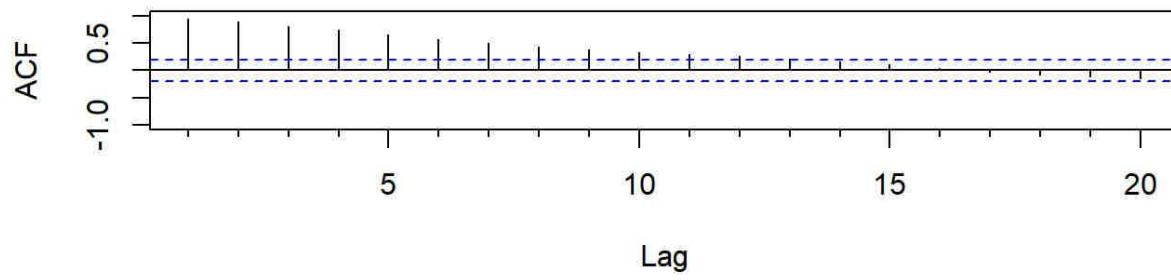
## 2 ANALISIS DE COINTEGRACION

### 2.1 Estacionariedad

```
library(forecast)
par(mfrow=c(2,1))
Acf(x=x,lag.max = 20, ylim=c(-1,1), xlim=c(1,20))
Pacf(x=y,lag.max = 20, ylim=c(-1,1),main="")
```



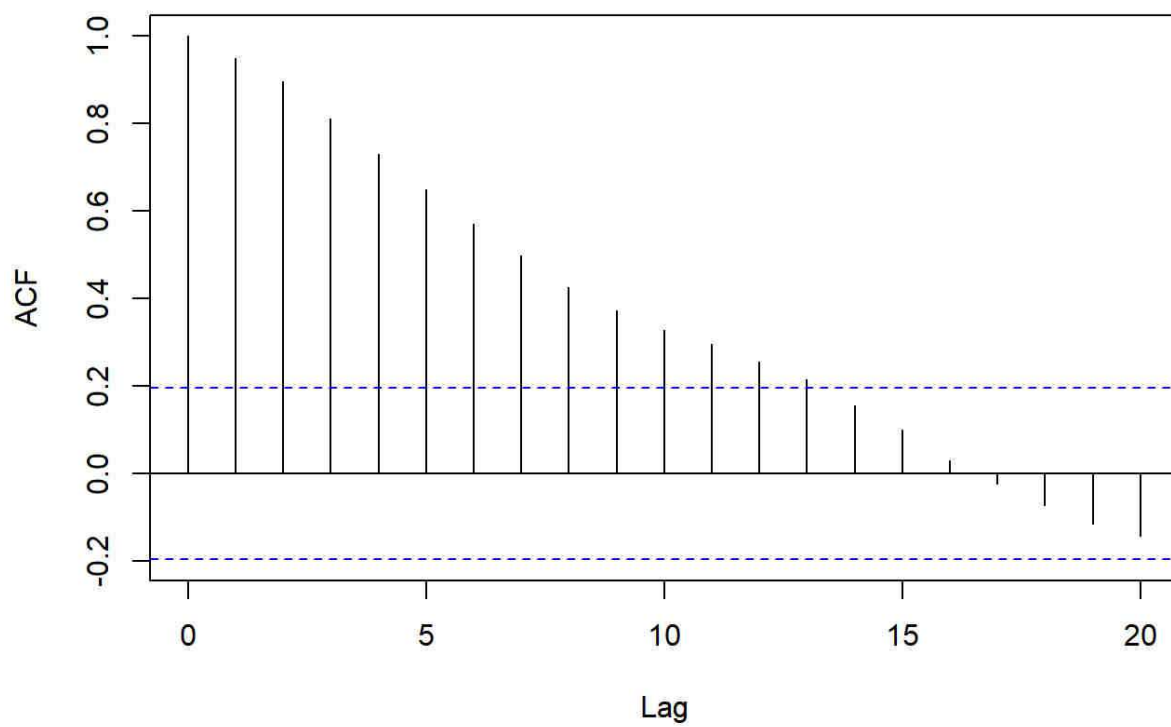
**Series x**



```
par(mfrow=c(1,1))
```

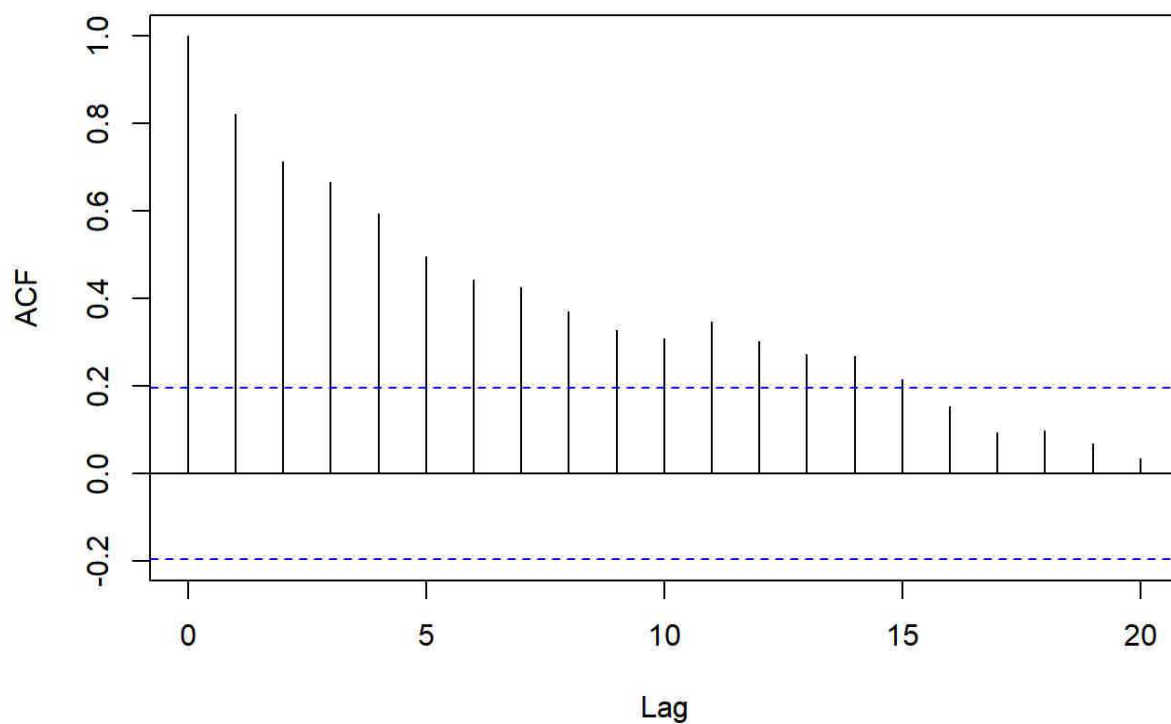
```
acf(x)
```

**Series x**

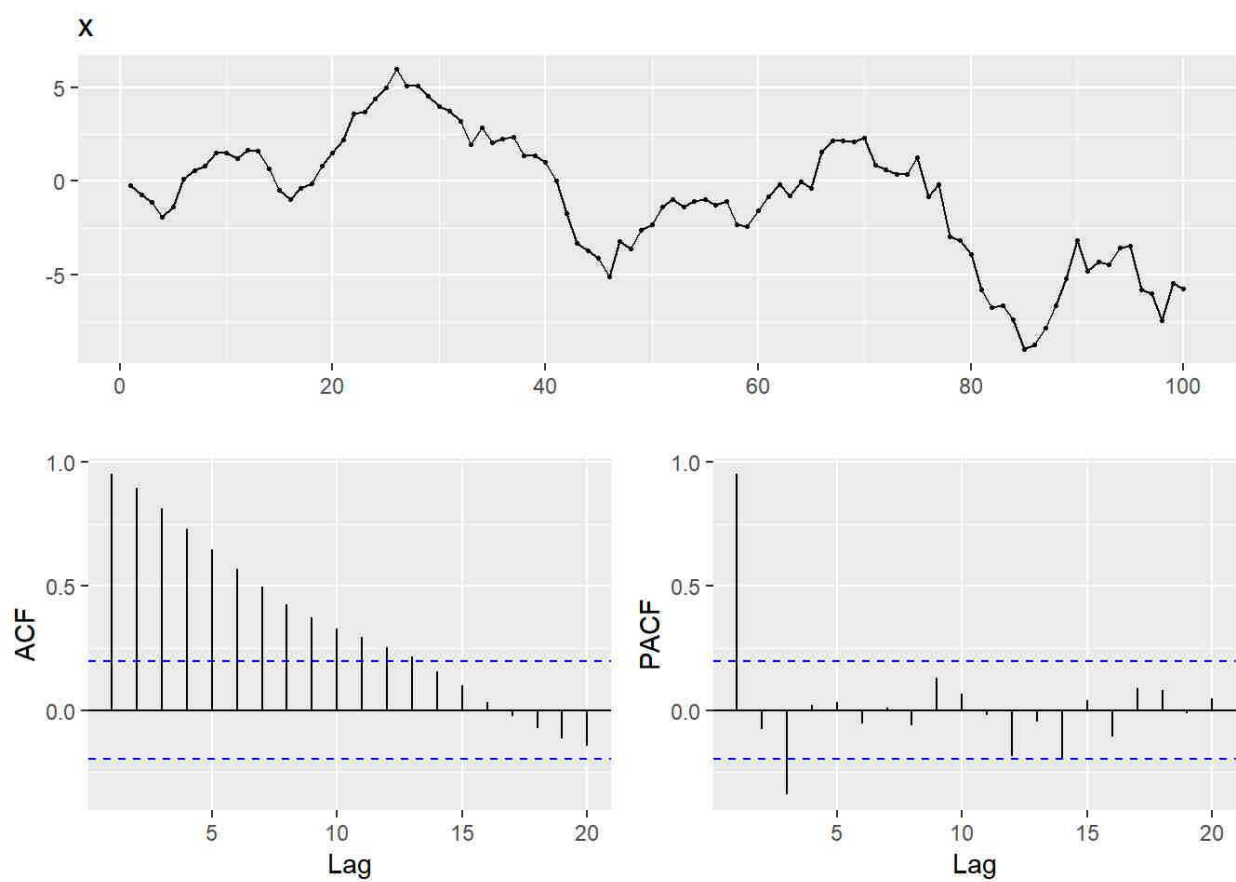


```
acf(y)
```

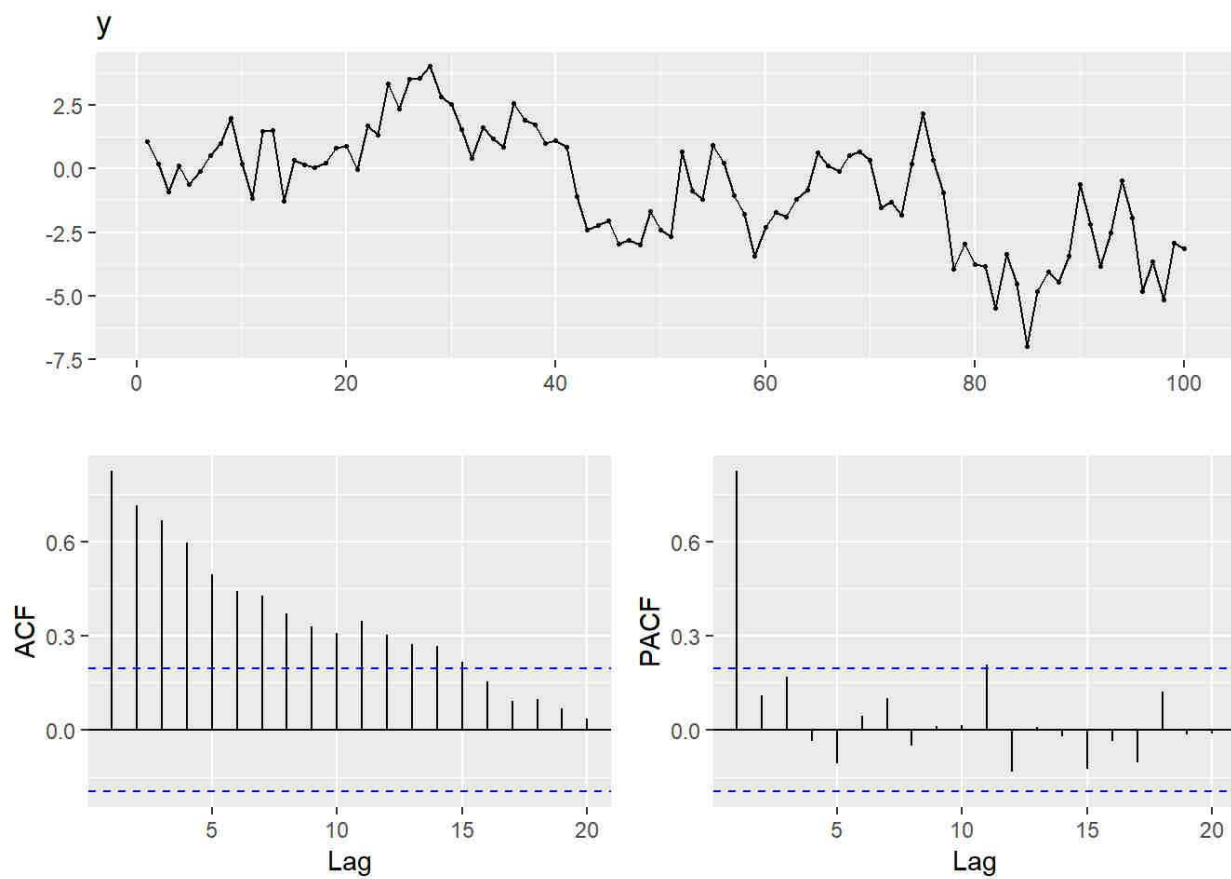
Series y



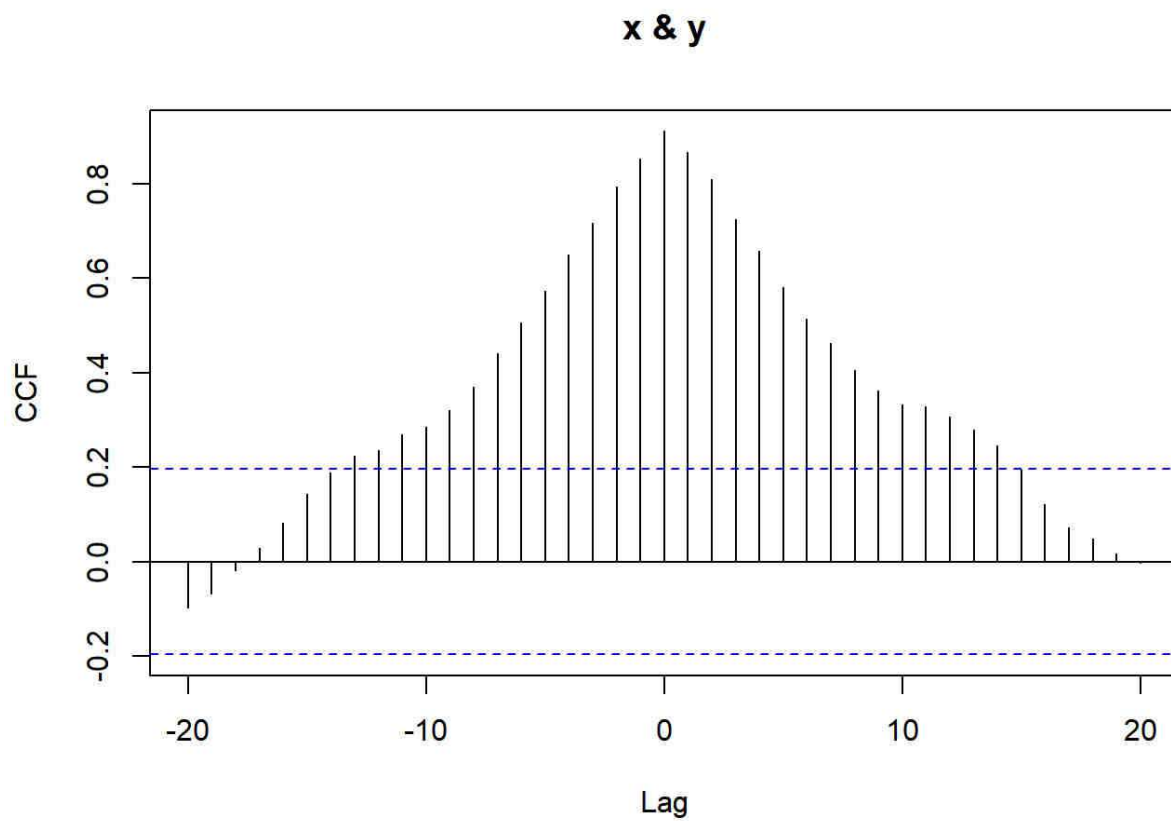
```
ggtsdisplay(x)
```



```
ggtsdisplay(y)
```



$\text{Ccf}(x,y)$



## 2.1.1 test de Raices Unitarias

```
library(tseries)
adf.test(na.omit(x))
```

```
##
## Augmented Dickey-Fuller Test
##
## data: na.omit(x)
## Dickey-Fuller = -3.1104, Lag order = 4, p-value = 0.1167
## alternative hypothesis: stationary
```

```
adf.test(na.omit(y))
```

```
##
## Augmented Dickey-Fuller Test
##
## data: na.omit(y)
## Dickey-Fuller = -3.3478, Lag order = 4, p-value = 0.06718
## alternative hypothesis: stationary
```

```
ndiffs(na.omit(x)) #Number of differences required to achieve stationarity (from the forecast packa
ge)
```

```
## [1] 1
```

```
ndiffs(na.omit(x)) #Number of differences required to achieve stationarity (from the forecast packa
ge)
```

```
## [1] 1
```

## 3 Test de Cointegración

### 3.1 Opción 1: Regresión sobre series diferenciadas para conseguir estacionariedad

```
# Opción uno Diferencio las series y compruebo que el modelo siga siendo válido

#construyo las series diferenciadas dy dx
dy <- diff(y)
dx <- diff(x)

dif.reg<-lm(dy~dx)
summary(dif.reg)
```

```
##
## Call:
## lm(formula = dy ~ dx)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2.06678 -0.79172 -0.08572  0.55367  3.03780
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -0.001199   0.113513  -0.011   0.992
## dx           0.744197   0.119427   6.231 1.19e-08 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.127 on 97 degrees of freedom
## Multiple R-squared:  0.2859, Adjusted R-squared:  0.2785
## F-statistic: 38.83 on 1 and 97 DF,  p-value: 1.194e-08
```

## 3.2 Opción 2: Test cointegración Engle-Granger

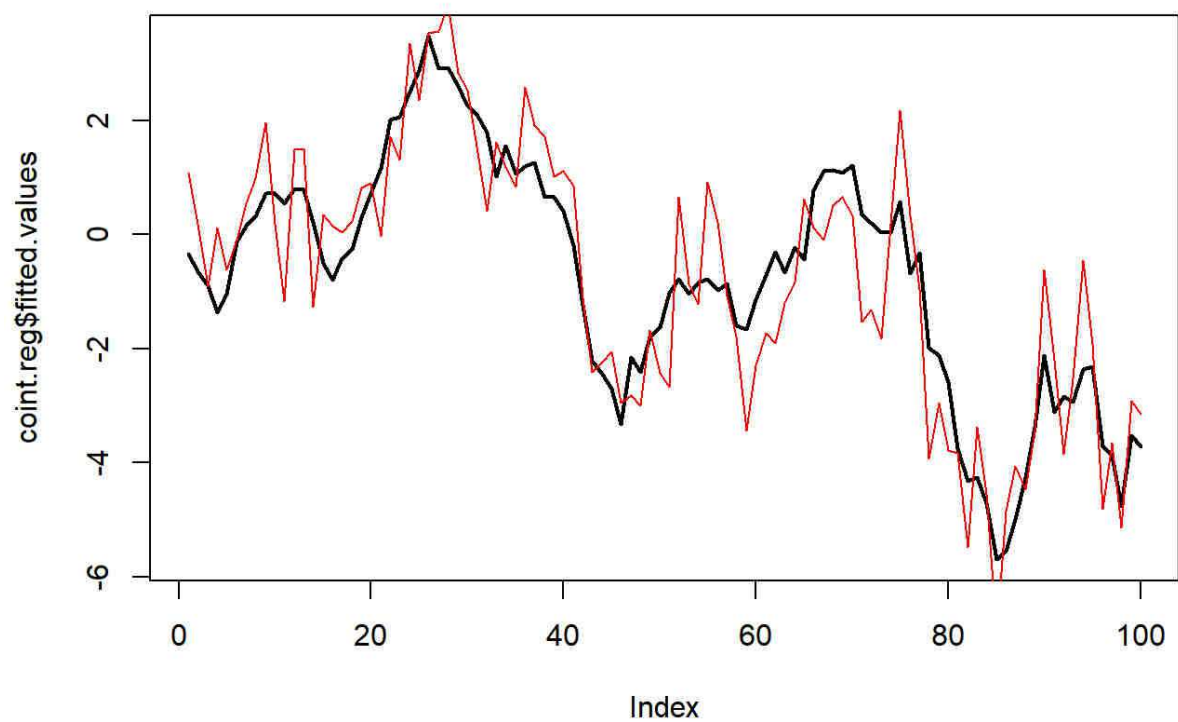
### 3.2.1 Primera etapa

```
#coint.reg <- lm(y ~ x -1)
coint.reg <- lm(y ~ x)

summary(coint.reg)
```

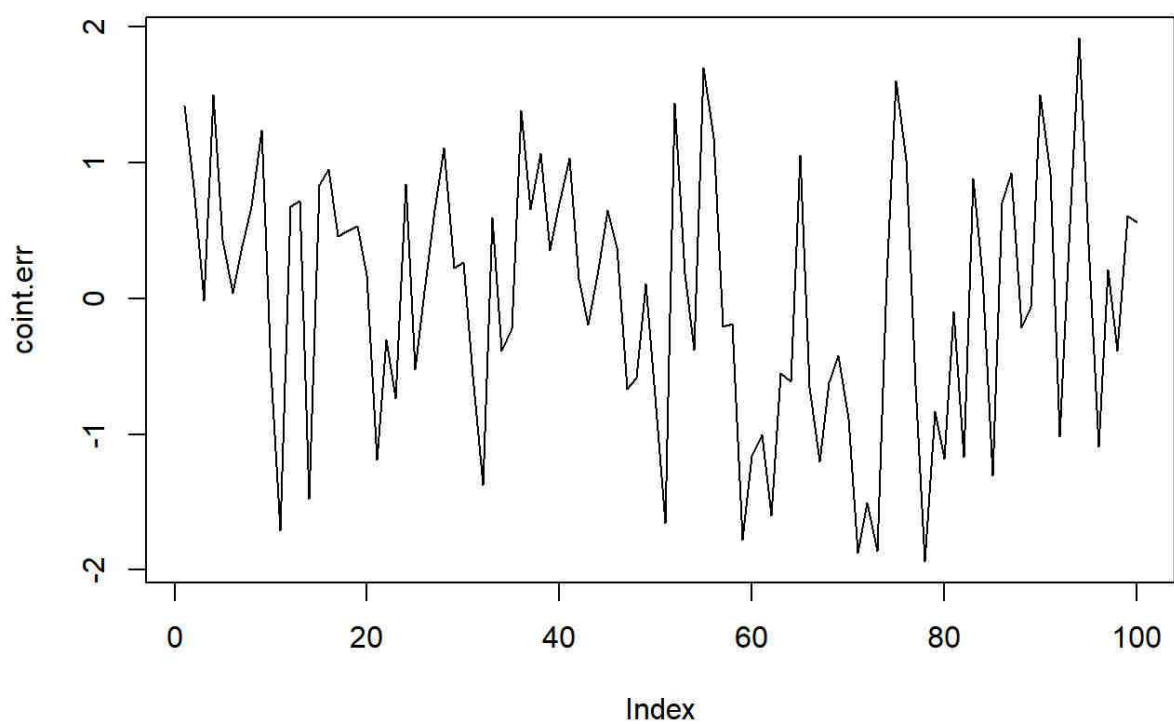
```
##
## Call:
## lm(formula = y ~ x)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.9367 -0.6431  0.1463  0.6817  1.9162
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -0.19151   0.09807  -1.953   0.0537 .
## x           0.61047   0.02799  21.808 <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.9467 on 98 degrees of freedom
## Multiple R-squared:  0.8291, Adjusted R-squared:  0.8274
## F-statistic: 475.6 on 1 and 98 DF,  p-value: < 2.2e-16
```

```
plot(coint.reg$fitted.values, type="l", lwd=2)
lines(y, type="l", col=2)
```



### 3.2.2 extraigo el error

```
# extraigo el error  
coint.err<-residuals(coint.reg)  
  
plot(coint.err, type="l")
```



### 3.2.3 segunda etapa: compruebo si el erroer es estacionario

```
# segunda etapa: compruebo si el erroer es estacionario
library(fUnitRoots)
```

```
## Warning: package 'fUnitRoots' was built under R version 3.4.4
```

```
## Loading required package: timeDate
```

```
## Loading required package: timeSeries
```

```
## Loading required package: fBasics
```

```
##
```

```
## Rmetrics Package fBasics
```

```
## Analysing Markets and calculating Basic Statistics
```

```
## Copyright (C) 2005-2014 Rmetrics Association Zurich
```

```
## Educational Software for Financial Engineering and Computational Science
```

```
## Rmetrics is free software and comes with ABSOLUTELY NO WARRANTY.
```

```
## https://www.rmetrics.org --- Mail to: info@rmetrics.org
```

```
adf.test(coint.err) # ojo el p-valor no es v?lido por eso tengo que calcular el p-valor de manera ro
busta:
```

```
## Warning in adf.test(coint.err): p-value smaller than printed p-value
```

```
##
## Augmented Dickey-Fuller Test
##
## data: coint.err
## Dickey-Fuller = -4.243, Lag order = 4, p-value = 0.01
## alternative hypothesis: stationary
```

```
punitroot(adf.test(coint.err)$statistic, trend="nc")
```

```
## Warning in adf.test(coint.err): p-value smaller than printed p-value
```

```
## Dickey-Fuller
## 2.335029e-05
```

```
library(urca)
```

```
## Warning: package 'urca' was built under R version 3.4.4
```

```
##  
## Attaching package: 'urca'
```

```
## The following objects are masked from 'package:fUnitRoots':  
##  
##      punitroot, qunitroot, unitrootTable
```

```
tru<-ur.df(coint.err, type = c("drift"), lags = 12,  
          selectlags = c("BIC"))  
summary(tru)
```

```
##  
## #####  
## # Augmented Dickey-Fuller Test Unit Root Test #  
## #####  
##  
## Test regression drift  
##  
##  
## Call:  
## lm(formula = z.diff ~ z.lag.1 + 1 + z.diff.lag)  
##  
## Residuals:  
##      Min       1Q   Median       3Q      Max   
## -1.68434 -0.72726 -0.00079  0.63470  1.92076   
##  
## Coefficients:  
##              Estimate Std. Error t value Pr(>|t|)      
## (Intercept) -0.05114    0.09786  -0.523   0.603      
## z.lag.1      -0.78296    0.13008  -6.019 4.43e-08 ***  
## z.diff.lag    0.08789    0.10851   0.810   0.420      
## ---  
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1  
##  
## Residual standard error: 0.9096 on 84 degrees of freedom  
## Multiple R-squared:  0.3656, Adjusted R-squared:  0.3505   
## F-statistic: 24.2 on 2 and 84 DF, p-value: 5.002e-09  
##  
##  
## Value of test-statistic is: -6.0193 18.1162  
##  
## Critical values for test statistics:  
##      1pct  5pct 10pct   
## tau2 -3.51 -2.89 -2.58   
## phi1  6.70  4.71  3.86
```

```
punitroot(tru@teststat[1], trend="c")
```

```
## [1] 1.076392e-07
```

### 3.2.4 Phillips–Ouliaris Cointegration Test HO: NO Cointegration



```
# Phillips-Ouliaris Cointegration Test H0: NO Cointegration

# null hypothesis that x is not cointegrated.
#series<-matrix(data=c(series,seriex),nrow=length(series),ncol=2)
seriey.ts<-ts(y, start=1, frequency=1)
seriex.ts<-ts(x, start=1, frequency=1)
series<-ts.intersect(seriey.ts,seriex.ts)

library(tseries)
po.test(series,demean = FALSE, lshort=TRUE) #Phillips-Ouliaris Cointegration Test H0: NO Cointegration
```

```
## Warning in po.test(series, demean = FALSE, lshort = TRUE): p-value smaller
## than printed p-value
```

```
##
## Phillips-Ouliaris Cointegration Test
##
## data: series
## Phillips-Ouliaris standard = -68.455, Truncation lag parameter =
## 0, p-value = 0.01
```

```
library(urca)
# null hypothesis that x is not cointegrated.
capocointest<-ca.po(series,demean=c("constant"),lag=c("short"))
summary(capocointest)
```

```
##
## #####
## # Phillips and Ouliaris Unit Root Test #
## #####
##
## Test of type Pu
## detrending of series with constant only
##
##
## Call:
## lm(formula = z[, 1] ~ z[, -1])
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.9367 -0.6431  0.1463  0.6817  1.9162
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -0.19151    0.09807  -1.953   0.0537 .
## z[, -1]      0.61047    0.02799  21.808  <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.9467 on 98 degrees of freedom
## Multiple R-squared:  0.8291, Adjusted R-squared:  0.8274
## F-statistic: 475.6 on 1 and 98 DF,  p-value: < 2.2e-16
##
##
## Value of test-statistic is: 84.4267
##
## Critical values of Pu are:
##              10pct    5pct    1pct
## critical values 27.8536 33.713 48.0021
```

```
# si el estadístico es mayor que en valor en tabla rechazamos la Ho nula
```

## 3.3 Opción 3: Test de Johansen

```
# Opción 3: Test de Johansen
# In Johansen cointegration test, the null hypothesis for
# the eigenvalue test is that there are cointegration relations.
# si el estadístico es mayor que en valor en tabla rechazamos la Ho nula

# The test is therefore sequential: you test first for r=0, then r=1, etc.
# The test concludes on the value of  when the test fails
# to reject  for the first time.

coint.test1 <- ca.jo(series,
                     ecdet = "trend",
                     type="trace",
                     K=2,
                     spec="longrun"
)
summary(coint.test1) # Check the cointegration rank here
```

```
##
## #####
## # Johansen-Procedure #
## #####
##
## Test type: trace statistic , with linear trend in cointegration
##
## Eigenvalues (lambda):
## [1] 3.457631e-01 5.624329e-02 1.632439e-18
##
## Values of teststatistic and critical values of test:
##
##          test 10pct  5pct  1pct
## r <= 1 |   5.67 10.49 12.25 16.26
## r = 0  |  47.25 22.76 25.32 30.45
##
## Eigenvectors, normalised to first column:
## (These are the cointegration relations)
##
##          series.ts.l2 seriex.ts.l2  trend.l2
## series.ts.l2   1.00000000   1.0000000  1.0000000
## seriex.ts.l2  -0.53339842  -4.4985061 -2.1945561
## trend.l2       0.01078375  -0.4397119  0.2395067
##
## Weights W:
## (This is the loading matrix)
##
##          series.ts.l2 seriex.ts.l2  trend.l2
## series.ts.d   -0.9987682  0.008830796 5.585293e-17
## seriex.ts.d   -0.2081063  0.019629749 1.356246e-17
```

```
coint.test1@cval
```

```
##          10pct  5pct  1pct
## r <= 1 |  10.49 12.25 16.26
## r = 0  |  22.76 25.32 30.45
```

```
coint.test1@teststat
```

```
## [1]  5.672913 47.252921
```

## 4 Modelo de Corrección del Error

```

# Una vez que sé que existe correlación puedo estimar el modelo de Corrección de Error

# contruyo la serie de residuos retardada
coint.err.lag<-coint.err[1:(length(coint.err)-1)]

#construyo las series diferenciadas dy dx
dy <- diff(y)
dx <- diff(x)

# construyo la serie dyt-1, dyt-1

dy.lag<-dy[1:(length(dy)-1)]
dx.lag <- dx[1:(length(dx)-1)]

# ajusto el tamaño de las series para que todas tenga el mismo tamaño

dy <- dy[2:length(dy)]
dx <- dx[2:length(dx)]
coint.err.lag<-coint.err.lag[2:length(coint.err.lag)]

#write.csv(data.frame(y,x,coint.err), file="clipboard")

#data.frame(dy,dx,coint.err.lag,dy.lag,dx.lag)

# Estimo el modelo de corrección de Error

ecm1 <- lm(dy ~ coint.err.lag + dy.lag + dx.lag)
summary(ecm1)

```

```

##
## Call:
## lm(formula = dy ~ coint.err.lag + dy.lag + dx.lag)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -3.2702 -0.6993 -0.0138  0.7562  2.6743
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)   -0.04593     0.11519   -0.399    0.691
## coint.err.lag -0.86769     0.15362  -5.648 1.72e-07 ***
## dy.lag         0.15793     0.12713    1.242    0.217
## dx.lag        -0.01662     0.14812   -0.112    0.911
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.138 on 94 degrees of freedom
## Multiple R-squared:  0.2915, Adjusted R-squared:  0.2689
## F-statistic: 12.89 on 3 and 94 DF, p-value: 3.981e-07

```

```

# otra alternativa es utilizar la librería "ecm" (Error Correction Models)
library(ecm)

```

```

## Warning: package 'ecm' was built under R version 3.4.4

```

```

# no la veremos aquí

# Existen otros test de Cointegración: el test de Johansen de cointegración
# la librería urca permite estimarlo

library(urca)

# Create a matrix of the cointegrated variables
cointvars <- cbind(y,x)
# Perform cointegration tests
coint.test1 <- ca.jo(cointvars,
                     ecdet = "const",
                     type="eigen",
                     K=2,
                     spec="longrun"
)
summary(coint.test1) # Check the cointegration rank here

```

```

##
## #####
## # Johansen-Procedure #
## #####
##
## Test type: maximal eigenvalue statistic (lambda max) , without linear trend and constant in coint
egration
##
## Eigenvalues (lambda):
## [1] 3.164790e-01 1.546551e-02 -2.874229e-17
##
## Values of teststatistic and critical values of test:
##
##          test 10pct 5pct 1pct
## r <= 1 | 1.53 7.52 9.24 12.97
## r = 0 | 37.29 13.75 15.67 20.20
##
## Eigenvectors, normalised to first column:
## (These are the cointegration relations)
##
##          y.l2      x.l2  constant
## y.l2      1.0000000  1.000000  1.000000
## x.l2     -0.5951140 -3.668918 -5.505703
## constant  0.2259113 -6.748026 34.147100
##
## Weights W:
## (This is the loading matrix)
##
##          y.l2      x.l2      constant
## y.d -0.8737578 0.006212134 -4.375262e-18
## x.d -0.1081495 0.010510345 -2.069909e-18

```

```

# In Johansen cointegration test, the null hypothesis for the eigenvalue test is that there are cointegration relations.
# The test is therefore sequential: you test first for  $\alpha$ , then  $\beta$ , etc.
# The test concludes on the value of  $\lambda$  when the test fails to reject  $H_0$  for the first time.

# Using the output of the test, estimate an ECM
ecm.res1 <- cajorls(coint.test1,
                    r = 1, # Cointegration rank
                    reg.number = 1) # which variable(s) to put on LHS
# (column indexes of cointvars)
summary(ecm.res1$rlm)

```

```

##
## Call:
## lm(formula = substitute(form1), data = data.mat)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -3.2955 -0.7317 -0.0294  0.7157  2.5729
##
## Coefficients:
##      Estimate Std. Error t value Pr(>|t|)
## ect1    -0.8738     0.1520  -5.748 1.09e-07 ***
## y.dl1   -0.7138     0.1280  -5.575 2.31e-07 ***
## x.dl1    0.5123     0.1509   3.394  0.001 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.129 on 95 degrees of freedom
## Multiple R-squared:  0.2959, Adjusted R-squared:  0.2737
## F-statistic: 13.31 on 3 and 95 DF, p-value: 2.525e-07

```



Proyecto de innovación  
“Innova Docencia” de la UCM  
Convocatoria 2018-2019  
Proyecto número 275

**R-adaptación de la asignatura de métodos econométricos en economía y finanzas del  
Grado de Estadística Aplicada**

Facultad de Estudios Estadísticos  
Departamento de Economía Aplicada, Pública y Política

**Título**

***Práctica 10: De vueltas con la ley de Okun: cointegración de las series  
de PIB y Empleo***

**Correspondiente al**

**Capítulo 4 MODELOS CON DATOS DE SERIES TEMPORALES I: COINTEGRACIÓN Y  
REGRESIONES ESPURIAS**

- Magia “potagia” o las regresiones espurias
- Cointegración y modelos de regresión con **series no estacionarias**
- ¿Cuántas veces hay que diferenciar una serie para convertirla en estacionaria? El contraste de raíces unitarias
- Contrastes de cointegración
- Cointegración y Mecanismos de Corrección del Error (MCE)
  - Práctica 8 : Abracadabra: No estacionariedad e media y Cointegración***
  - Práctica 9 : Cointegración y Modelos de Corrección del Error***
  - Práctica 10: De vueltas con la ley de Okun: cointegración de las series de PIB y Empleo***

# De vueltas con la Ley de Okun: Cointegración de las series de PIB y empleo

*Asignatura*  
*Métodos Econométricos en Economía y Finanzas*

*Profesores*  
*Loren Escot, Alicia Pérez-Alonso y Julio E. Sandubete*

Dpto Economía Aplicada, Pública y Política  
Facultad Estudios Estadísticos - Universidad Complutense de Madrid

*Curso académico 2018/2019*

## **Abstract**

En esta práctica seguimos con el análisis de *cointegración* aplicándolo en esta ocasión a la ya conocida ley de Okun. Se analizará el orden de integración, la cointegración y el modelo de corrección del error entre las series de PIB y Empleo generadas en prácticas previas.

El contenido de esta práctica corresponde con la tercera tarea solicitada a los alumnos en la evaluación continua.

- 1 Tarea 3. ¿Están Cointegradas las series de PIB y EMPLEO?
- 2 Carga de los datos de PIB y Empleo
- 3 Regresión de la serie en niveles
- 4 Análisis de la Estacionariedad en media
- 5 Análisis de Cointegración
- 6 Modelo de Corrección del error
- 7 Posibles ampliaciones

## 1 Tarea 3. ¿Están Cointegradas las series de PIB y EMPLEO?

Teneis que utilizar las series de PIB (índice de Volumen encadenado) y Empleo (puestos de trabajo equivalente a tiempo completo) para intentar averiguar si estas series están cointegradas.

En caso afirmativo intentad estimar un modelo de corrección del error. # Carga de librerías necesarias

```
#=====#
# File: P3OkunCoint.R
# This script was written for R.3.4.3
# First version: 23abr2018
# Last version: 04abr2019
# Last revisor: Loren Escot
#_____#
# Description: Práctica 3, ¿están cointegradas las series de PIB y Empleo?
#
# Input: okundat.Rdata (PIB / Empleo Equivalente)
#=====#
dev.off() # Cierra gráficos que podamos tener abiertos
```



```
## null device
##      1
```

```
rm(list=ls())          # Elimina todos los objetos que tengamos en la memoria

if(!require(forecast)){install.packages("forecast")}
```

```
## Loading required package: forecast
```

```
if(!require(seasonal)){install.packages("seasonal")}
```

```
## Loading required package: seasonal
```

```
## Warning: package 'seasonal' was built under R version 3.4.1
```

```
if(!require(lmtest)){install.packages("lmtest")}
```

```
## Loading required package: lmtest
```

```
## Warning: package 'lmtest' was built under R version 3.4.2
```

```
## Loading required package: zoo
```

```
##
## Attaching package: 'zoo'
```

```
## The following objects are masked from 'package:base':
##
##      as.Date, as.Date.numeric
```

```
if(!require(tseries)){install.packages("tseries")}
```

```
## Loading required package: tseries
```

```
if(!require(urca)){install.packages("urca")}
```

```
## Loading required package: urca
```

```
## Warning: package 'urca' was built under R version 3.4.4
```

```
if(!require(data.table)){install.packages("data.table")}
if(!require(dynlm)){install.packages("dynlm")}
```

```
## Loading required package: dynlm
```

```
## Warning: package 'dynlm' was built under R version 3.4.4
```

```
if(!require(stargazer)){install.packages("stargazer")}
```

```
## Loading required package: stargazer
```

```
## Warning: package 'stargazer' was built under R version 3.4.1
```

```
##  
## Please cite as:
```

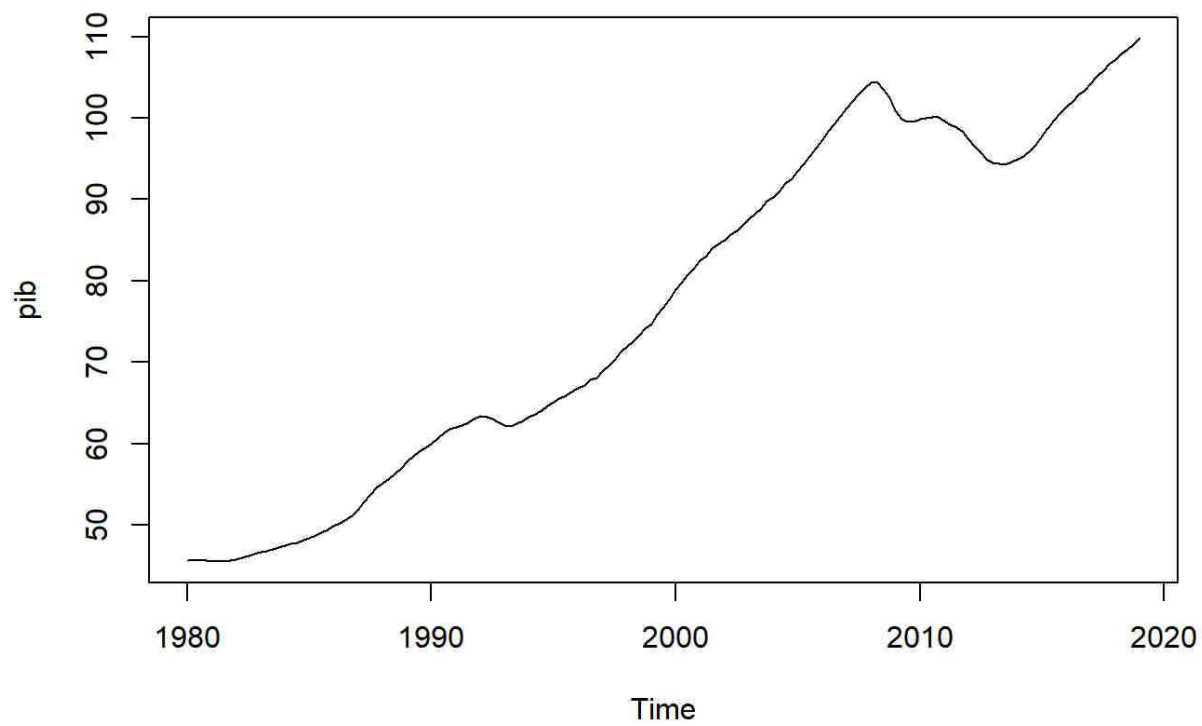
```
## Hlavac, Marek (2015). stargazer: Well-Formatted Regression and Summary Statistics Tables.
```

```
## R package version 5.2. http://CRAN.R-project.org/package=stargazer
```

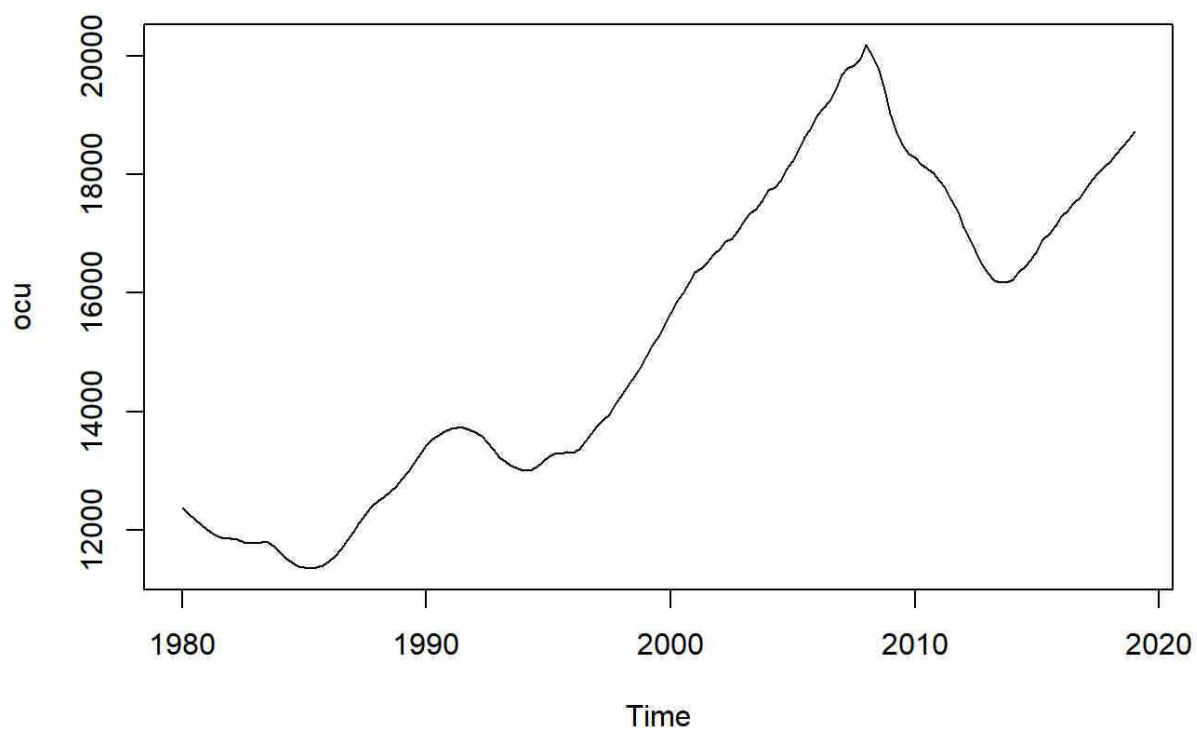
```
library(forecast)           # ggtdisplay(); ggAcf(); ggPacf()  
library(seasonal)          # trend()  
library(lmtest)            # dwtest()  
library(tseries)          # adf.test()  
library(urca)              # ur.test()  
#library(data.table)       # shift() - Fast Lead/Lag For Vectors And Lists  
library(dynlm)              # lm con objetos serie temporal ts (equidistantes) o zoo (no necesar  
iamente equidistantes)  
library(stargazer)  
#_____#
```

## 2 Carga de los datos de PIB y Empleo

```
# Datos  
#getwd()  
#setw("yourpathher")  
load(file="okundat.Rdata")  
  
pib <- window(okun[,2], start=1980)  
ocu <- window(okun[,3], start=1980)  
datos <- ts.intersect(pib, ocu, dframe = TRUE)  
  
# Inspección gráfica de las series  
# Serie creciente - no revierte a una media (cte)  
# Sospecho que no son estacionarias  
plot(pib)
```



```
plot(ocu)
```



```
# Logaritmo de las series en niveles (no estacionarias?) - tienen escalas distintas
lpib0 <- log(pib)
locu0 <- log(ocu)

# Para asegurarme de que no tiene componente estacional (lo ajusto estacionalmente)
# ...los datos de partidas son desestacionalizados y sin efecto de calendario
lpib <- trend(seas(lpib0))
```

```
## Model used in SEATS is different: (0 2 1)
```

```
locu <- trend(seas(locu0))

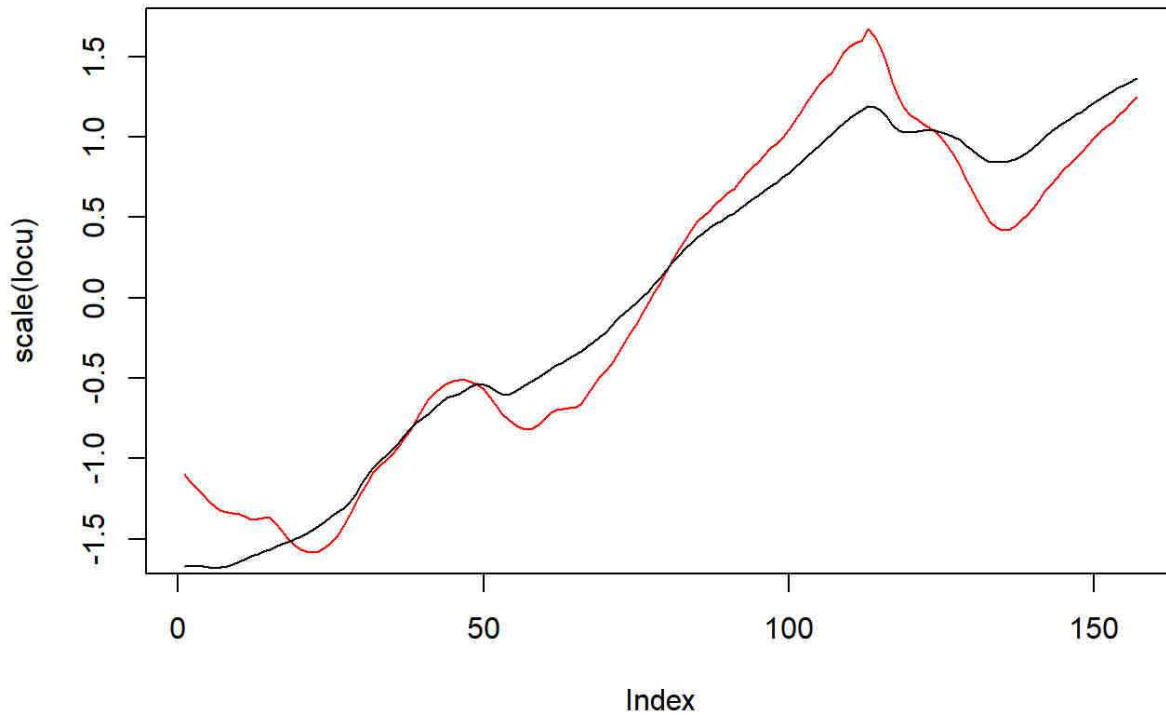
# ¿Son muy diferentes las series?
rbind(lpib0,lpib)[,1:10]
```

```
##           [,1]      [,2]      [,3]      [,4]      [,5]      [,6]      [,7]
## lpib0 3.820187 3.820687 3.820910 3.820055 3.818713 3.817949 3.818752
## lpib  3.820192 3.820719 3.820848 3.820046 3.818711 3.817991 3.818759
##           [,8]      [,9]     [,10]
## lpib0 3.821081 3.824609 3.829130
## lpib  3.821071 3.824635 3.829118
```

```
rbind(locu0,locu)[,1:10]
```

```
##           [,1]      [,2]      [,3]      [,4]      [,5]      [,6]      [,7]
## locu0 9.423253 9.415522 9.407814 9.401018 9.39448 9.388225 9.383315
## locu  9.423251 9.415467 9.407913 9.400974 9.39441 9.388210 9.383434
##           [,8]      [,9]     [,10]
## locu0 9.381061 9.381044 9.379320
## locu  9.381289 9.380775 9.379159
```

```
plot(scale(locu), type="l", col=2)
lines(scale(lpib), type="l")
```



```
# _____ #
```

### 3 Regresión de la serie en niveles

```
# Regresión de la serie en niveles
modelo.lm <- lm(locu~lpib)
summary(modelo.lm)
```

```
##
## Call:
## lm(formula = locu ~ lpib)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.071068 -0.035477 -0.008451  0.031949  0.091984
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  7.07328    0.05230  135.25  <2e-16 ***
## lpib         0.59107    0.01212   48.75  <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.04383 on 155 degrees of freedom
## Multiple R-squared:  0.9388, Adjusted R-squared:  0.9384
## F-statistic: 2377 on 1 and 155 DF, p-value: < 2.2e-16
```

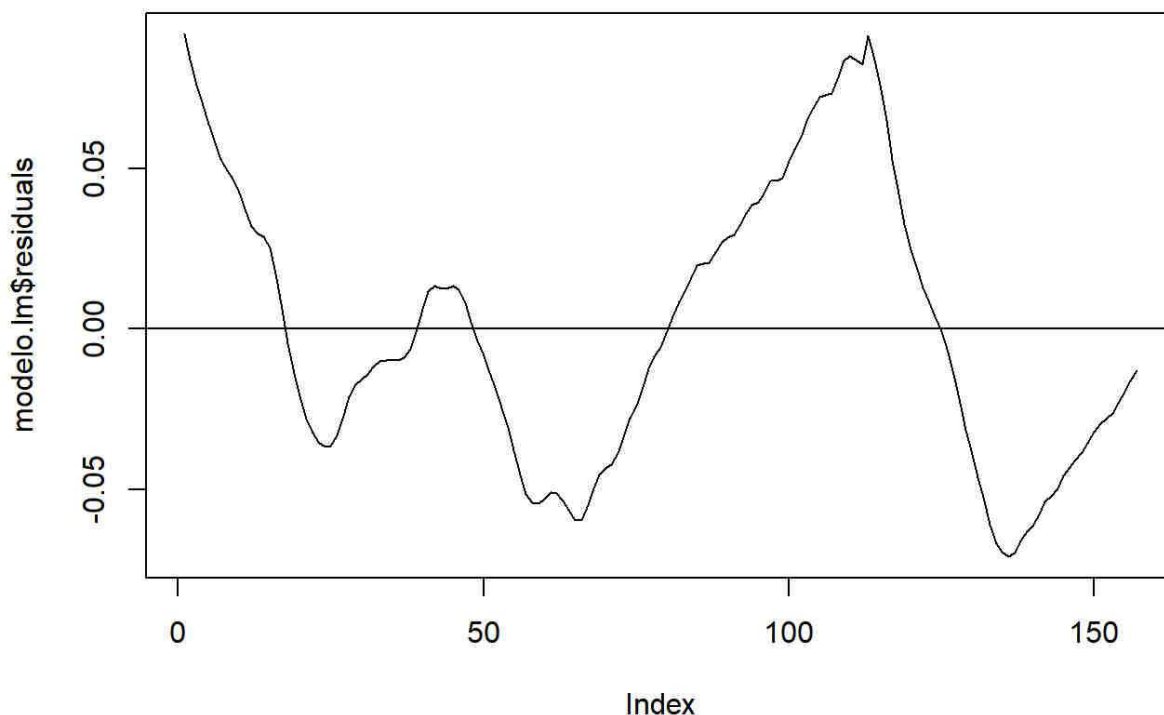
```
# Síntomas relación espuria
# 1. Significatividad (muy alta) de los parámetros
# 2. R2 muy alto
# 3. Residuos presentan autocorrelación (DW próximo a cero - autocorrelación 1er orden positiva)

# Completad vosotros: análisis gráfico residuos visto para Normalidad

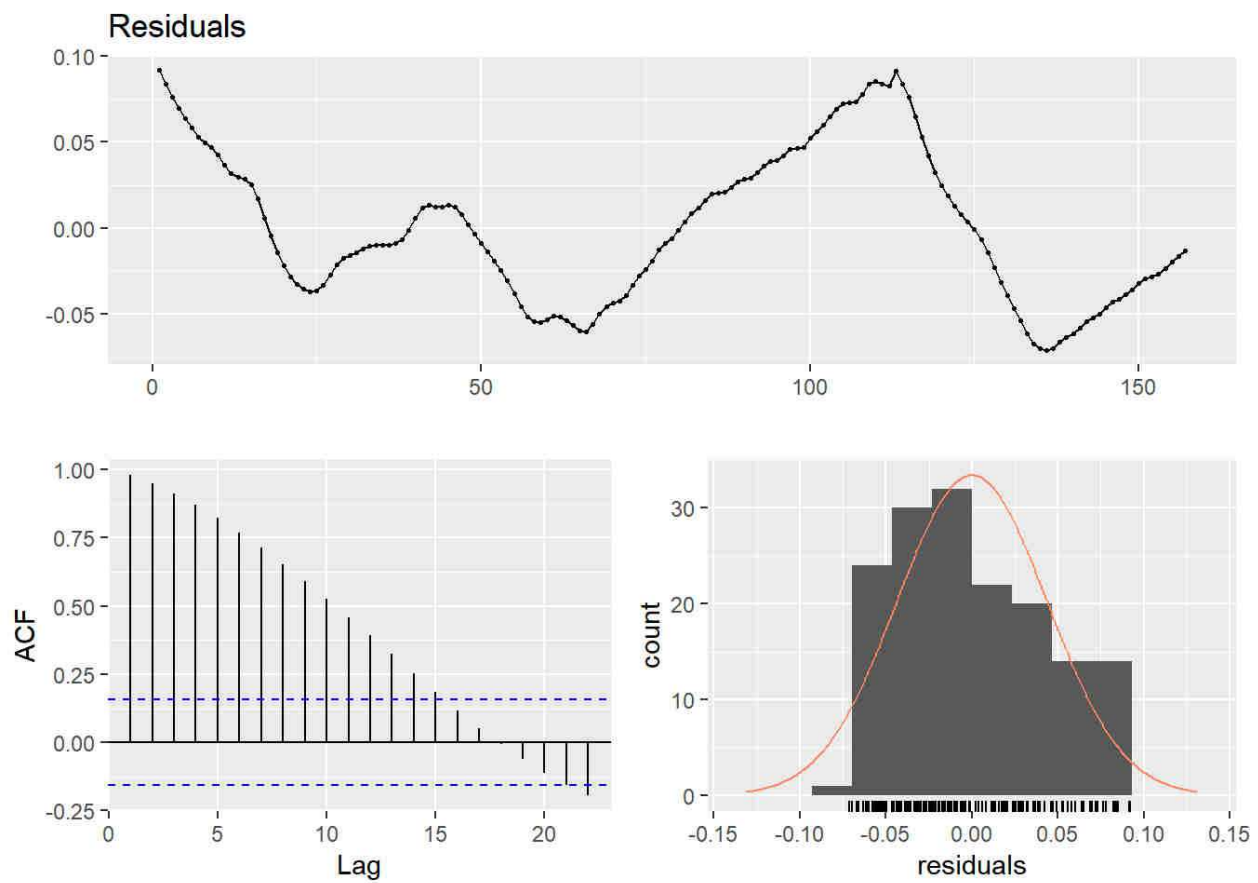
# Test de Durbin-Watson
# H0: No autocorrelación
# H1: Autocorrelación de orden 1 (positiva)
dwtest(modelo.lm)
```

```
##
## Durbin-Watson test
##
## data: modelo.lm
## DW = 0.012454, p-value < 2.2e-16
## alternative hypothesis: true autocorrelation is greater than 0
```

```
# Análisis gráfico - CORRELOGRAMA
# Los residuos NO parecen un RUIDO BLANCO
plot(modelo.lm$residuals, type="l")
abline(h=0)
```

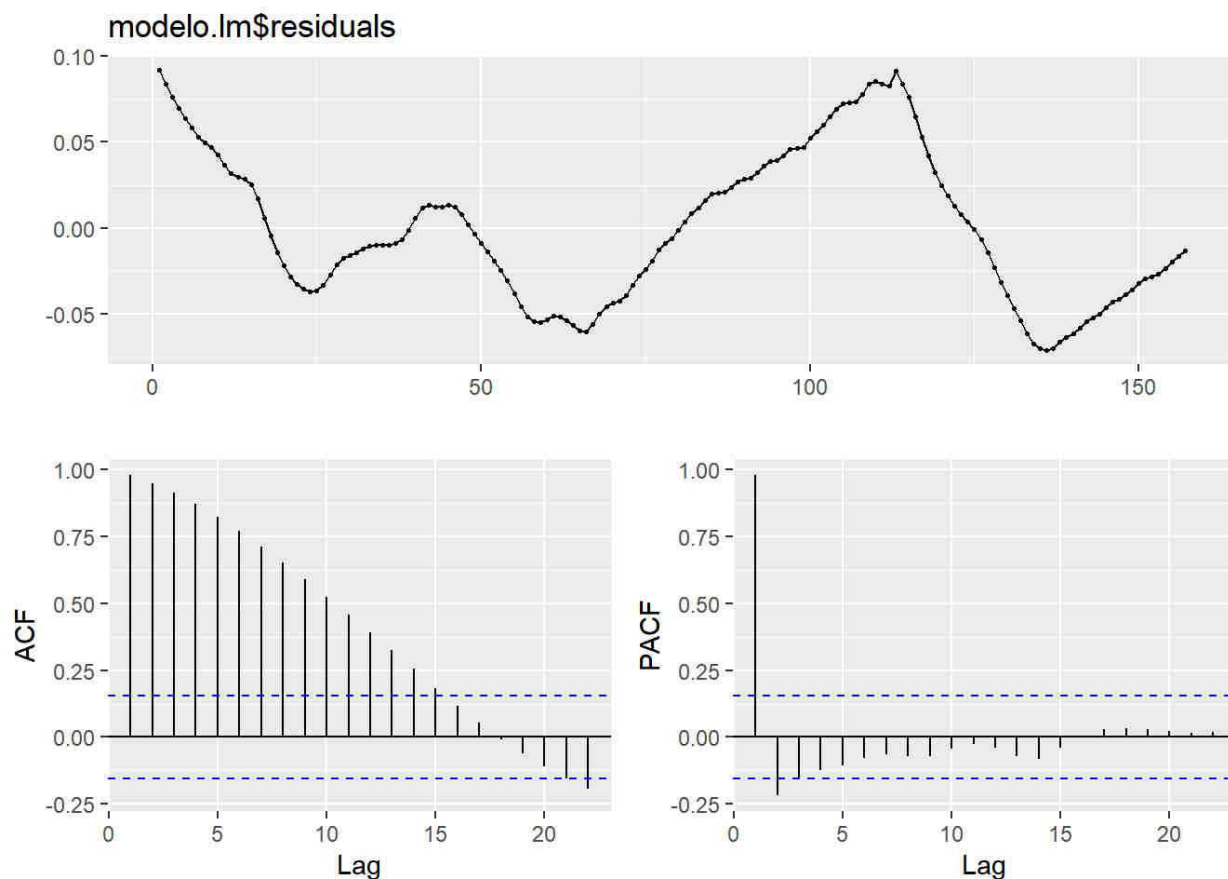


```
# Gráfico residuos, acf, histograma
# Test de Ljung-Box (LB)
# No se restringe a autocorrelación de orden 1 (orden "m")
# H0: ausencia de autocorrelación de orden p
# H0: Las autocorrelaciones hasta orden p son iguales a cero
# H0: Los datos son aleatorios e independientes (hasta orden p)
checkresiduals(modelo.lm, test="LB")
```



```
##
##  Ljung-Box test
##
## data:  residuals
## Q* = 1029.1, df = 8, p-value < 2.2e-16
##
## Model df: 2.   Total lags used: 10
```

```
# Gráfico residuos, acf, pacf
ggtsdisplay(modelo.lm$residuals) # Objeto ggplot
```



## 4 Análisis de la Estacionariedad en media

```
# _____#
# MUY IMPORTANTE:
#Correlación no implica no estacionariedad, piensa en un MA(1).
#La no estacionariedad tiene que ver con la presencia de raíces unitarias.
#Correlogramas:
#PACF: AR(1), la particularidad es que esa autocorrelación está muy próxima a 1.
#ACF: el parámetro del primer retardo está muy próximo a 1 y decae lentamente a cero.
#Todo esto nos hace sospechar que podemos estar ante una raíz unitaria.
# _____#
# ¿Son las series estacionarias? ADF test
#Una serie es estacionaria si sus propiedades estadísticas (momentos de primer y segundo orden)
#permanecen constantes a lo largo del tiempo
adf.test(na.omit(lpib))
```

```
##
## Augmented Dickey-Fuller Test
##
## data: na.omit(lpib)
## Dickey-Fuller = -2.9088, Lag order = 5, p-value = 0.1975
## alternative hypothesis: stationary
```

```
adf.test(na.omit(locu))
```



```
##
## Augmented Dickey-Fuller Test
##
## data: na.omit(locu)
## Dickey-Fuller = -2.3324, Lag order = 5, p-value = 0.4378
## alternative hypothesis: stationary
```

```
# En ambos casos aceptamos H0: (raíz unitaria)

# Number of differences required to achieve stationarity (from the forecast package)
# Valores predeterminados para implementar ADF
# ndiffs(x#, alpha = 0.05, test = c("kpss", "adf", "pp"), max.d = 2)
ndiffs(na.omit(lpib)) # PIB ~ I(2) 2 raíces unitarias
```

```
## [1] 2
```

```
ndiffs(na.omit(locu)) # ocu ~ I(1) 1 raíz unitaria
```

```
## [1] 1
```

```
ndiffs(na.omit(locu), test="adf") # ocu ~ I(1) 1 raíz unitaria
```

```
## [1] 2
```

```
# En base a este primer análisis preliminar, PIB Y ocu NO
# pueden estar cointegradas porque tienen órdenes de integración distintos.

# Augmented dickey-Fuller (ADF) test
# ur.df() ==> type = c("none", "drift", "trend")
# Interpretación:
# https://stats.stackexchange.com/questions/24072/interpreting-rs-ur-df-dickey-fuller-unit-root-test-results
# _____#
# 1) type="none":  $\Delta y(t) = \gamma * y(t-1) + e(t)$ 
# tau1: Ho:  $\gamma = 0$  (unit root)

# 2) type="drift":  $\Delta y(t) = a_0 + \gamma * y(t-1) + e(t)$ 
# tau2: Ho:  $\gamma=0$  (unit root).
# phi1: combined Ho:  $\gamma = a_0 = 0$  (unit root and no drift). Rejecting this null implies that one OR two was NOT zero.

# 3) type="trend":  $\Delta y(t) = a_0 + \gamma * y(t-1) + a_2(t) + e(t)$ 
# tau3: Ho:  $\gamma=0$  (unit root)
# phi2: Ho:  $\gamma = a_0 = a_2 = 0$ . (unit root, no drift term AND no time trend term). Rejecting this null implies that one, two, OR all three of these terms was NOT zero
# Se corresponde con  $a_0$ 
# phi3:  $\gamma = a_2 = 0$  (unit root AND no trend at same time). Rejecting this null implies that one OR two was NOT zero.
# En un modelo con drift, contraste si necesito o no tendencia:  $a_2$ 
# _____#
# SERIE PIB EN NIVELES
# Paso 1: Estimo modelo completo con drift y trend
tru <- ur.df(lpib, type = c("trend"), lags = 12,
             selectlags = c("BIC"))
summary(tru)
```

```
##
## #####
## # Augmented Dickey-Fuller Test Unit Root Test #
## #####
##
## Test regression trend
##
##
## Call:
## lm(formula = z.diff ~ z.lag.1 + 1 + tt + z.diff.lag)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.0032736 -0.0006538 -0.0000797  0.0007436  0.0035516
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  1.249e-02  6.885e-03   1.814 0.072059 .
## z.lag.1      -3.148e-03  1.829e-03  -1.721 0.087573 .
## tt           1.718e-05  1.187e-05   1.447 0.150368
## z.diff.lag1   2.210e+00  8.530e-02  25.912 < 2e-16 ***
## z.diff.lag2  -2.627e+00  2.043e-01 -12.858 < 2e-16 ***
## z.diff.lag3   2.923e+00  2.935e-01   9.961 < 2e-16 ***
## z.diff.lag4  -3.436e+00  3.640e-01  -9.439 < 2e-16 ***
## z.diff.lag5   3.701e+00  4.334e-01   8.538 3.26e-14 ***
## z.diff.lag6  -3.324e+00  4.882e-01  -6.809 3.35e-10 ***
## z.diff.lag7   2.725e+00  4.891e-01   5.572 1.40e-07 ***
## z.diff.lag8  -2.208e+00  4.364e-01  -5.059 1.42e-06 ***
## z.diff.lag9   1.665e+00  3.675e-01   4.531 1.32e-05 ***
## z.diff.lag10 -1.126e+00  2.980e-01  -3.778 0.000241 ***
## z.diff.lag11  6.856e-01  2.074e-01   3.306 0.001227 **
## z.diff.lag12 -2.335e-01  8.732e-02  -2.674 0.008466 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.001134 on 129 degrees of freedom
## Multiple R-squared:  0.9669, Adjusted R-squared:  0.9633
## F-statistic: 269.3 on 14 and 129 DF, p-value: < 2.2e-16
##
##
## Value of test-statistic is: -1.7214 3.1742 2.1559
##
## Critical values for test statistics:
##      1pct  5pct 10pct
## tau3 -3.99 -3.43 -3.13
## phi2  6.22  4.75  4.07
## phi3  8.43  6.49  5.47
```

```
# Me fijo en tau3 (H0: gamma=0)
# Acepto H0 (a todos los niveles de significación), implica que tengo una raíz unitaria
# phi2: H0: gamma=drift=trend=0 => No rechazo H0
# phi3: H0: gamma=trend=0 => No rechazo H0; NO NECESITO TENDENCIA EN MODELO CON CONSTANTE

# Paso 2: En base al resultado anterior, estimodo modelo con drift pero sin trend
tru <- ur.df(lpib, type = c("drift"), lags = 12,
             selectlags = c("BIC"))
summary(tru)
```

```
##
## #####
## # Augmented Dickey-Fuller Test Unit Root Test #
## #####
##
## Test regression drift
##
##
## Call:
## lm(formula = z.diff ~ z.lag.1 + 1 + z.diff.lag)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.0034192 -0.0006588 -0.0000202  0.0006367  0.0035442
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)   0.0028204   0.0016705    1.688 0.093750 .
## z.lag.1       -0.0005583   0.0003764   -1.483 0.140426
## z.diff.lag1    2.2314836   0.0843768   26.447 < 2e-16 ***
## z.diff.lag2   -2.6647264   0.2034550  -13.097 < 2e-16 ***
## z.diff.lag3    2.9684291   0.2930072   10.131 < 2e-16 ***
## z.diff.lag4   -3.4969906   0.3630412   -9.632 < 2e-16 ***
## z.diff.lag5    3.7711551   0.4324973    8.719 1.14e-14 ***
## z.diff.lag6   -3.4063727   0.4869391   -6.995 1.25e-10 ***
## z.diff.lag7    2.8096362   0.4875896    5.762 5.71e-08 ***
## z.diff.lag8   -2.2900252   0.4344920   -5.271 5.50e-07 ***
## z.diff.lag9    1.7280985   0.3664514    4.716 6.12e-06 ***
## z.diff.lag10  -1.1801638   0.2968597   -3.975 0.000116 ***
## z.diff.lag11   0.7311254   0.2058398    3.552 0.000533 ***
## z.diff.lag12  -0.2681287   0.0843238   -3.180 0.001843 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.001139 on 130 degrees of freedom
## Multiple R-squared:  0.9664, Adjusted R-squared:  0.963
## F-statistic: 287.5 on 13 and 130 DF, p-value: < 2.2e-16
##
##
## Value of test-statistic is: -1.4833 3.6836
##
## Critical values for test statistics:
##      1pct  5pct 10pct
## tau2 -3.46 -2.88 -2.57
## phi1  6.52  4.63  3.81
```

```
# Me fijo en tau2 (H0: gamma=0)
# Acepto H0 (a todos los niveles de significación), implica que tengo una raíz unitaria
# Phi1: H0: gamma=0 y cte=0 => No rechazo H0

# Paso 3: En base al resultado anterior, estimo un modelo sin drift y sin trend
tru <- ur.df(lpib, type = c("none"), lags = 12,
             selectlags = c("BIC"))
summary(tru)
```

```
##
## #####
## # Augmented Dickey-Fuller Test Unit Root Test #
## #####
##
## Test regression none
##
##
## Call:
## lm(formula = z.diff ~ z.lag.1 - 1 + z.diff.lag)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.0035934 -0.0006752 -0.0000325  0.0006848  0.0035363
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## z.lag.1          7.443e-05  3.527e-05   2.110 0.036723 *
## z.diff.lag1      2.251e+00  8.419e-02  26.736 < 2e-16 ***
## z.diff.lag2     -2.689e+00  2.044e-01 -13.158 < 2e-16 ***
## z.diff.lag3      2.993e+00  2.947e-01  10.157 < 2e-16 ***
## z.diff.lag4     -3.524e+00  3.652e-01  -9.648 < 2e-16 ***
## z.diff.lag5      3.805e+00  4.351e-01   8.745 9.39e-15 ***
## z.diff.lag6     -3.437e+00  4.900e-01  -7.015 1.11e-10 ***
## z.diff.lag7      2.833e+00  4.908e-01   5.772 5.38e-08 ***
## z.diff.lag8     -2.308e+00  4.374e-01  -5.275 5.33e-07 ***
## z.diff.lag9      1.744e+00  3.689e-01   4.727 5.79e-06 ***
## z.diff.lag10    -1.191e+00  2.989e-01  -3.985 0.000112 ***
## z.diff.lag11     7.366e-01  2.073e-01   3.554 0.000528 ***
## z.diff.lag12    -2.696e-01  8.491e-02  -3.176 0.001865 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.001147 on 131 degrees of freedom
## Multiple R-squared:  0.9829, Adjusted R-squared:  0.9813
## F-statistic: 580.9 on 13 and 131 DF, p-value: < 2.2e-16
##
##
## Value of test-statistic is: 2.1104
##
## Critical values for test statistics:
##      1pct  5pct 10pct
## tau1 -2.58 -1.95 -1.62
```

```
# Me fijo en tau1 (H0: gamma=0)
# Acepto H0 (a todos los niveles de significación), implica que tengo una raíz unitaria
# _____#
# SERIE OCUPADOS EN NIVELES
# Paso 1: Modelo completo
tru <- ur.df(locu, type = c("trend"), lags = 12,
             selectlags = c("BIC"))
summary(tru)
```

```
##
## #####
## # Augmented Dickey-Fuller Test Unit Root Test #
## #####
##
## Test regression trend
##
##
## Call:
## lm(formula = z.diff ~ z.lag.1 + 1 + tt + z.diff.lag)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.0194558 -0.0011907  0.0001071  0.0012203  0.0097756
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  7.053e-02  2.685e-02   2.627  0.00958 **
## z.lag.1      -7.524e-03  2.879e-03  -2.614  0.00995 **
## tt           2.634e-05  1.172e-05   2.247  0.02621 *
## z.diff.lag1  1.126e+00  8.227e-02  13.681 < 2e-16 ***
## z.diff.lag2 -1.969e-01  8.311e-02  -2.369  0.01923 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.002627 on 139 degrees of freedom
## Multiple R-squared:  0.8895, Adjusted R-squared:  0.8863
## F-statistic: 279.6 on 4 and 139 DF,  p-value: < 2.2e-16
##
##
## Value of test-statistic is: -2.6135 2.7794 3.4331
##
## Critical values for test statistics:
##      1pct  5pct 10pct
## tau3 -3.99 -3.43 -3.13
## phi2  6.22  4.75  4.07
## phi3  8.43  6.49  5.47
```

```
# Paso 2: Modelo seleccionado en base a phi3
tru <- ur.df(locu, type = c("drift"), lags = 12,
             selectlags = c("BIC"))
summary(tru)
```

```
##
## #####
## # Augmented Dickey-Fuller Test Unit Root Test #
## #####
##
## Test regression drift
##
##
## Call:
## lm(formula = z.diff ~ z.lag.1 + 1 + z.diff.lag)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.0206175 -0.0011349  0.0002047  0.0013204  0.0089800
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  0.017161   0.012700   1.351  0.17879
## z.lag.1      -0.001751   0.001318  -1.329  0.18605
## z.diff.lag1  1.162845   0.081734  14.227 < 2e-16 ***
## z.diff.lag2 -0.244127   0.081555  -2.993  0.00326 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.002664 on 140 degrees of freedom
## Multiple R-squared:  0.8854, Adjusted R-squared:  0.883
## F-statistic: 360.7 on 3 and 140 DF,  p-value: < 2.2e-16
##
##
## Value of test-statistic is: -1.3289 1.5983
##
## Critical values for test statistics:
##      1pct  5pct 10pct
## tau2 -3.46 -2.88 -2.57
## phi1  6.52  4.63  3.81
```

```
# Paso 3: Modelo seleccionado en base a phi1
# También en base a phi2 paso 1
tru <- ur.df(locu, type = c("none"), lags = 12,
             selectlags = c("BIC"))
summary(tru)
```

```
##
## #####
## # Augmented Dickey-Fuller Test Unit Root Test #
## #####
##
## Test regression none
##
##
## Call:
## lm(formula = z.diff ~ z.lag.1 - 1 + z.diff.lag)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.0212309 -0.0011933  0.0002385  0.0013576  0.0085150
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## z.lag.1      2.917e-05  2.498e-05   1.167  0.24503
## z.diff.lag1  1.178e+00  8.124e-02  14.496 < 2e-16 ***
## z.diff.lag2 -2.575e-01  8.119e-02  -3.171  0.00186 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.002672 on 141 degrees of freedom
## Multiple R-squared:  0.901, Adjusted R-squared:  0.8989
## F-statistic: 427.6 on 3 and 141 DF, p-value: < 2.2e-16
##
##
## Value of test-statistic is: 1.1674
##
## Critical values for test statistics:
##      1pct  5pct 10pct
## tau1 -2.58 -1.95 -1.62
```

# Conclusión: En base a tau1, aceptamos  $H_0$  ( $\gamma=0$ ), esto es, hay una raíz unitaria

```
# _____#
# Hemos encontrado una raíz unitaria, ¿La serie en diferencias es estacionaria?
# _____#
# SERIE PIB EN PRIMERSAS DIFERENCIAS
# Paso 1: Modelo con drift y trend
tru <- ur.df(diff(lplib,differences=1), type = c("trend"), lags = 12,
             selectlags = c("BIC"))
summary(tru)
```

```
##
## #####
## # Augmented Dickey-Fuller Test Unit Root Test #
## #####
##
## Test regression trend
##
##
## Call:
## lm(formula = z.diff ~ z.lag.1 + 1 + tt + z.diff.lag)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.0034782 -0.0006366  0.0000047  0.0006256  0.0035363
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)   6.512e-04  3.035e-04   2.145 0.033812 *
## z.lag.1       -6.886e-02  2.315e-02  -2.974 0.003510 **
## tt            -2.884e-06  2.502e-06  -1.153 0.251108
## z.diff.lag1    1.307e+00  8.311e-02  15.727 < 2e-16 ***
## z.diff.lag2   -1.368e+00  1.344e-01 -10.179 < 2e-16 ***
## z.diff.lag3    1.612e+00  1.715e-01   9.400 2.64e-16 ***
## z.diff.lag4   -1.900e+00  2.023e-01  -9.391 2.77e-16 ***
## z.diff.lag5    1.889e+00  2.400e-01   7.871 1.25e-12 ***
## z.diff.lag6   -1.537e+00  2.564e-01  -5.994 1.92e-08 ***
## z.diff.lag7    1.292e+00  2.414e-01   5.354 3.81e-07 ***
## z.diff.lag8   -1.016e+00  2.028e-01  -5.010 1.75e-06 ***
## z.diff.lag9    7.262e-01  1.740e-01   4.174 5.46e-05 ***
## z.diff.lag10  -4.660e-01  1.342e-01  -3.472 0.000704 ***
## z.diff.lag11   2.749e-01  8.503e-02   3.233 0.001554 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.001147 on 129 degrees of freedom
## Multiple R-squared:  0.7268, Adjusted R-squared:  0.6993
## F-statistic: 26.4 on 13 and 129 DF, p-value: < 2.2e-16
##
##
## Value of test-statistic is: -2.974 2.9533 4.4245
##
## Critical values for test statistics:
##      1pct  5pct 10pct
## tau3 -3.99 -3.43 -3.13
## phi2  6.22  4.75  4.07
## phi3  8.43  6.49  5.47
```

```
# Me fijo en tau3 (H0: gamma=0)
# Acepto H0 (a todos los niveles de significación), implica que tengo una raíz unitaria

# Paso 2: Modelo seleccionado en base a phi3
tru <- ur.df(diff(lpib,differences=1), type = c("drift"), lags = 12,
             selectlags = c("BIC"))
summary(tru)
```



```
##
## #####
## # Augmented Dickey-Fuller Test Unit Root Test #
## #####
##
## Test regression drift
##
##
## Call:
## lm(formula = z.diff ~ z.lag.1 + 1 + z.diff.lag)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.0035760 -0.0006771 -0.0000342  0.0006685  0.0035149
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)   0.0003520   0.0001576    2.233 0.027259 *
## z.lag.1       -0.0590859   0.0215736   -2.739 0.007032 **
## z.diff.lag1    1.3059328   0.0832102   15.694 < 2e-16 ***
## z.diff.lag2   -1.3783073   0.1342872  -10.264 < 2e-16 ***
## z.diff.lag3    1.6099557   0.1717314    9.375 2.87e-16 ***
## z.diff.lag4   -1.9090592   0.2023691   -9.434 < 2e-16 ***
## z.diff.lag5    1.8898676   0.2403342    7.863 1.26e-12 ***
## z.diff.lag6   -1.5424552   0.2566435   -6.010 1.75e-08 ***
## z.diff.lag7    1.2875550   0.2416221    5.329 4.23e-07 ***
## z.diff.lag8   -1.0185267   0.2030013   -5.017 1.69e-06 ***
## z.diff.lag9    0.7242196   0.1741984    4.157 5.80e-05 ***
## z.diff.lag10  -0.4663370   0.1343856   -3.470 0.000706 ***
## z.diff.lag11   0.2706687   0.0850542    3.182 0.001828 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.001148 on 130 degrees of freedom
## Multiple R-squared:  0.724, Adjusted R-squared:  0.6985
## F-statistic: 28.42 on 12 and 130 DF, p-value: < 2.2e-16
##
##
## Value of test-statistic is: -2.7388 3.7559
##
## Critical values for test statistics:
##      1pct  5pct 10pct
## tau2 -3.46 -2.88 -2.57
## phi1  6.52  4.63  3.81
```

```
# Paso 3: Modelo seleccionado en base a phi1
# También en base a phi2 de paso 1
tru <- ur.df(diff(lpib,differences=1), type = c("none"), lags = 12,
             selectlags = c("BIC"))
summary(tru)
```

```
##
## #####
## # Augmented Dickey-Fuller Test Unit Root Test #
## #####
##
## Test regression none
##
##
## Call:
## lm(formula = z.diff ~ z.lag.1 - 1 + z.diff.lag)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.0035289 -0.0004597  0.0000710  0.0007676  0.0038776
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## z.lag.1      -0.02092    0.01336  -1.566 0.119865
## z.diff.lag1   1.29734    0.08438  15.376 < 2e-16 ***
## z.diff.lag2  -1.42058    0.13495 -10.526 < 2e-16 ***
## z.diff.lag3   1.60035    0.17427   9.183 8.02e-16 ***
## z.diff.lag4  -1.94618    0.20473  -9.506 < 2e-16 ***
## z.diff.lag5   1.88764    0.24396   7.737 2.40e-12 ***
## z.diff.lag6  -1.56549    0.26031  -6.014 1.69e-08 ***
## z.diff.lag7   1.26445    0.24505   5.160 8.92e-07 ***
## z.diff.lag8  -1.02545    0.20604  -4.977 1.99e-06 ***
## z.diff.lag9   0.70467    0.17661   3.990 0.000109 ***
## z.diff.lag10 -0.46497    0.13641  -3.409 0.000868 ***
## z.diff.lag11  0.24714    0.08567   2.885 0.004583 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.001166 on 131 degrees of freedom
## Multiple R-squared:  0.7134, Adjusted R-squared:  0.6872
## F-statistic: 27.18 on 12 and 131 DF, p-value: < 2.2e-16
##
##
## Value of test-statistic is: -1.5656
##
## Critical values for test statistics:
##      1pct  5pct 10pct
## tau1 -2.58 -1.95 -1.62
```

```
# Me fijo en tau1 (H0: gamma=0)
# Acepto H0 (a todos los niveles de significación), implica que tengo una raíz unitaria
```

```
# _____#
# SERIE OCUPADOS EN PRIMERSAS DIFERENCIAS
# Paso 1: Modelo con drift y trend
tru <- ur.df(diff(locu,differences=1), type = c("trend"), lags = 12,
             selectlags = c("BIC"))
summary(tru)
```

```
##
## #####
## # Augmented Dickey-Fuller Test Unit Root Test #
## #####
##
## Test regression trend
##
##
## Call:
## lm(formula = z.diff ~ z.lag.1 + 1 + tt + z.diff.lag)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.0211871 -0.0011810  0.0002292  0.0013405  0.0085444
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  3.578e-04  5.292e-04   0.676  0.50007
## z.lag.1      -8.042e-02  2.954e-02  -2.722  0.00732 **
## tt           -8.716e-07  5.479e-06  -0.159  0.87383
## z.diff.lag    2.565e-01  8.198e-02   3.128  0.00214 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.00269 on 139 degrees of freedom
## Multiple R-squared:  0.0951, Adjusted R-squared:  0.07557
## F-statistic: 4.87 on 3 and 139 DF, p-value: 0.002988
##
##
## Value of test-statistic is: -2.7219 2.4803 3.7117
##
## Critical values for test statistics:
##      1pct  5pct 10pct
## tau3 -3.99 -3.43 -3.13
## phi2  6.22  4.75  4.07
## phi3  8.43  6.49  5.47
```

```
# Me fijo en tau3 (H0: gamma=0)
# Acepto H0 (a todos los niveles de significación), implica que tengo una raíz unitaria

# Paso 2: Modelo seleccionado en base a phi3
tru <- ur.df(diff(locu,differences=1), type = c("drift"), lags = 12,
             selectlags = c("BIC"))
summary(tru)
```

```
##
## #####
## # Augmented Dickey-Fuller Test Unit Root Test #
## #####
##
## Test regression drift
##
##
## Call:
## lm(formula = z.diff ~ z.lag.1 + 1 + z.diff.lag)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.0212134 -0.0011970  0.0002281  0.0013582  0.0085207
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  0.0002831  0.0002426   1.167  0.24539
## z.lag.1      -0.0799337  0.0292856  -2.729  0.00716 **
## z.diff.lag    0.2562145  0.0816806   3.137  0.00208 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.002681 on 140 degrees of freedom
## Multiple R-squared:  0.09494,    Adjusted R-squared:  0.08201
## F-statistic: 7.343 on 2 and 140 DF,  p-value: 0.0009279
##
##
## Value of test-statistic is: -2.7295 3.7338
##
## Critical values for test statistics:
##      1pct  5pct 10pct
## tau2 -3.46 -2.88 -2.57
## phi1  6.52  4.63  3.81
```

```
# Paso 3: Modelo seleccionado en base a phi1
# También en base a phi2 de paso 1
tru <- ur.df(diff(locu,differences=1), type = c("none"), lags = 12,
             selectlags = c("BIC"))
summary(tru)
```

```
##
## #####
## # Augmented Dickey-Fuller Test Unit Root Test #
## #####
##
## Test regression none
##
##
## Call:
## lm(formula = z.diff ~ z.lag.1 - 1 + z.diff.lag)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.0210426 -0.0009338  0.0004516  0.0015643  0.0087581
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## z.lag.1    -0.06687    0.02710  -2.468  0.01478 *
## z.diff.lag  0.25107    0.08167   3.074  0.00253 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.002684 on 141 degrees of freedom
## Multiple R-squared:  0.08633,    Adjusted R-squared:  0.07337
## F-statistic: 6.661 on 2 and 141 DF,  p-value: 0.001721
##
##
## Value of test-statistic is: -2.4681
##
## Critical values for test statistics:
##      1pct  5pct 10pct
## tau1 -2.58 -1.95 -1.62
```

```
# Dependiendo del nivel de significación escogido, aceptamos o rechazamos H0
# Soy conservadora (escojo 1%), de acuerdo a tau1 al 1%, acepto H0
# Si escojo 5% ó 10%, rechazo H0, no hay otra raíz unitaria
# y eso implicaría que OCU es I(1)
# Si PIB es I(2) y OCU es I(1) las series no están cointegradas, tienen que analizarse
# de forma independiente

# OJO: Importancia de especificar bien el modelo ADF.
# Resultado modelo con "trend" distinto a modelo "none"

# _____#
# Como he encontrado una raíz unitaria en la serie diferenciada, paso a diferenciar la serie dos veces
# _____#
# SERIE PIB diferenciada dos veces
tru <- ur.df(diff(lpib,differences=2), type = c("trend"), lags = 12,
             selectlags = c("BIC"))
summary(tru)
```

```

##
## #####
## # Augmented Dickey-Fuller Test Unit Root Test #
## #####
##
## Test regression trend
##
##
## Call:
## lm(formula = z.diff ~ z.lag.1 + 1 + tt + z.diff.lag)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.0033035 -0.0007262 -0.0000196  0.0007410  0.0035418
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  2.771e-05  2.229e-04   0.124 0.901295
## z.lag.1      -6.512e-01  1.600e-01  -4.070 8.16e-05 ***
## tt          -1.688e-07  2.394e-06  -0.071 0.943882
## z.diff.lag1   9.849e-01  1.579e-01   6.238 5.97e-09 ***
## z.diff.lag2  -5.437e-01  1.605e-01  -3.389 0.000934 ***
## z.diff.lag3   1.176e+00  1.665e-01   7.064 9.24e-11 ***
## z.diff.lag4  -9.812e-01  1.664e-01  -5.897 3.10e-08 ***
## z.diff.lag5   1.135e+00  1.867e-01   6.079 1.29e-08 ***
## z.diff.lag6  -7.415e-01  1.734e-01  -4.277 3.68e-05 ***
## z.diff.lag7   8.672e-01  1.803e-01   4.810 4.16e-06 ***
## z.diff.lag8  -5.425e-01  1.324e-01  -4.097 7.37e-05 ***
## z.diff.lag9   4.581e-01  1.328e-01   3.449 0.000762 ***
## z.diff.lag10 -2.901e-01  8.923e-02  -3.251 0.001470 **
## z.diff.lag11  1.969e-01  8.752e-02   2.250 0.026146 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.001167 on 128 degrees of freedom
## Multiple R-squared:  0.6994, Adjusted R-squared:  0.6689
## F-statistic: 22.91 on 13 and 128 DF, p-value: < 2.2e-16
##
##
## Value of test-statistic is: -4.0705 5.5316 8.2946
##
## Critical values for test statistics:
##      1pct  5pct 10pct
## tau3 -3.99 -3.43 -3.13
## phi2  6.22  4.75  4.07
## phi3  8.43  6.49  5.47

```

```

# Me fijo en tau3 (H0: gamma=0), rechazo H0 (no hay raíz unitaria)
# Hemos encontrado que el PIB es una serie I(2)
# Necesito diferenciar la serie dos veces para que sea estacionaria
# phi2: Rechazo que gamma=trend=drift=0 al 5% y al 10% (acepto al 1%)
# phi3: Rechazo que gamma=trend=0 al 5% y al 10% (acepto al 1%)

# Tengo que escoger nivel de significación para especificar el modelo
# Si arriba escogí 1% debería ser consistente con esa elección

# COMPLETAD VOSOTROS
# _____#
# SERIE OCUPADOS diferenciada dos veces
tru <- ur.df(diff(locu,differences=2), type = c("trend"), lags = 12,
             selectlags = c("BIC"))
summary(tru)

```

```

##
## #####
## # Augmented Dickey-Fuller Test Unit Root Test #
## #####
##
## Test regression trend
##
##
## Call:
## lm(formula = z.diff ~ z.lag.1 + 1 + tt + z.diff.lag)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.0215980 -0.0012654  0.0001738  0.0013798  0.0085028
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  5.519e-05  5.261e-04   0.105   0.917
## z.lag.1      -7.740e-01  1.062e-01 -7.288 2.21e-11 ***
## tt           -9.363e-08  5.656e-06 -0.017   0.987
## z.diff.lag   -8.629e-03  8.480e-02 -0.102   0.919
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.002762 on 138 degrees of freedom
## Multiple R-squared:  0.3915, Adjusted R-squared:  0.3783
## F-statistic: 29.6 on 3 and 138 DF, p-value: 7.709e-15
##
##
## Value of test-statistic is: -7.2877 17.7064 26.5573
##
## Critical values for test statistics:
##      1pct  5pct 10pct
## tau3 -3.99 -3.43 -3.13
## phi2  6.22  4.75  4.07
## phi3  8.43  6.49  5.47

```

```

# Me fijo en tau3 (H0: gamma=0), rechazo H0 (no hay una raíz unitaria)
# Hemos encontrado que OCU es una serie I(2)
# Necesito diferenciar la serie dos veces para que sea estacionaria

# phi2: Rechazo que gamma=trend=drift=0 a todos los niveles de significación
# phi3: Rechazo que gamma=trend=0 al 5% y al 10% a todos los niveles de significación

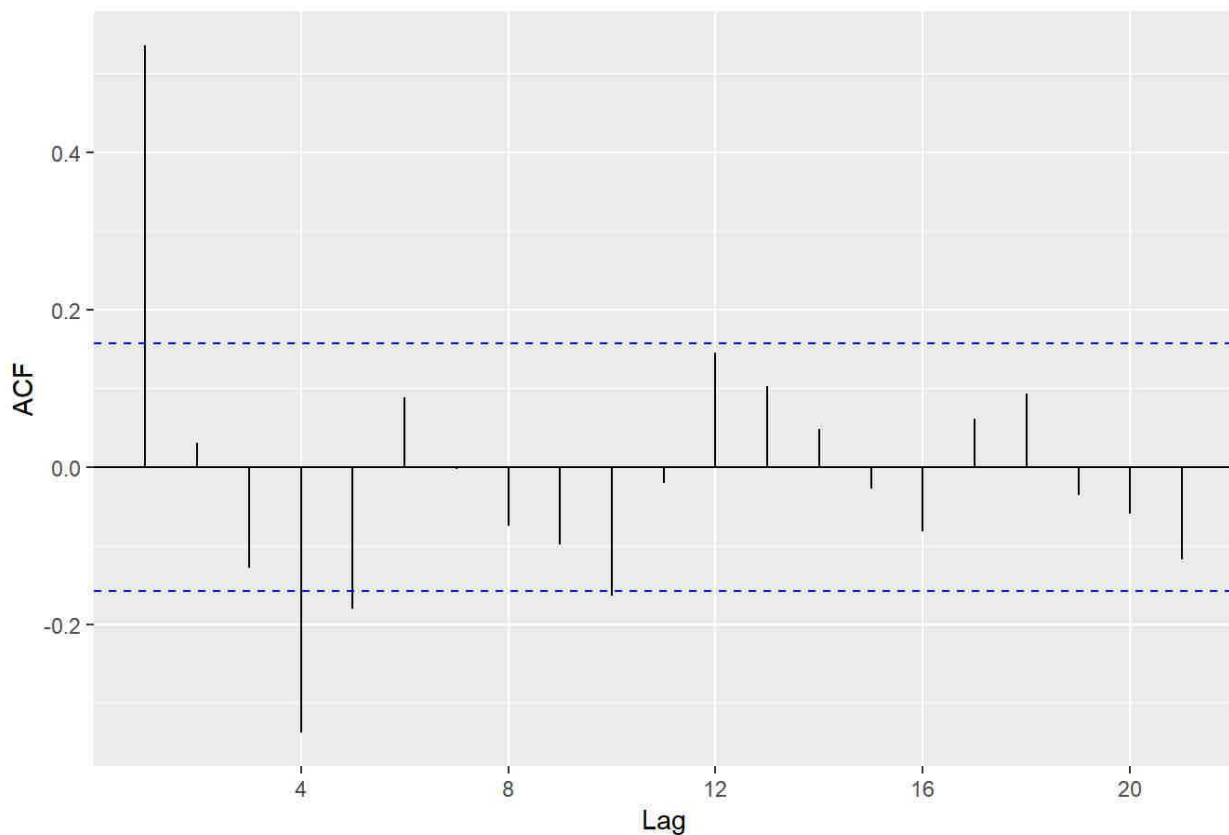
# Puedo considerar éste como mi modelo final

# _____#
# Series diferenciadas dos veces (tasa de aceleración)
#Log(yt)-log(yt-1)
#Es la tasa logarítmica de variación de una variable.
#Es un indicador de crecimiento relativo.
#Si se multiplica por 100 es la tasa de crecimiento porcentual de la variable.
#D(Log(yt)-log(yt-1))
#Es el cambio en la tasa logarítmica de variación de una variable.
#Es un indicador de la aceleración de la tasa de crecimiento relativo de una variable.

ggAcf(diff(lpib, differences = 2))

```

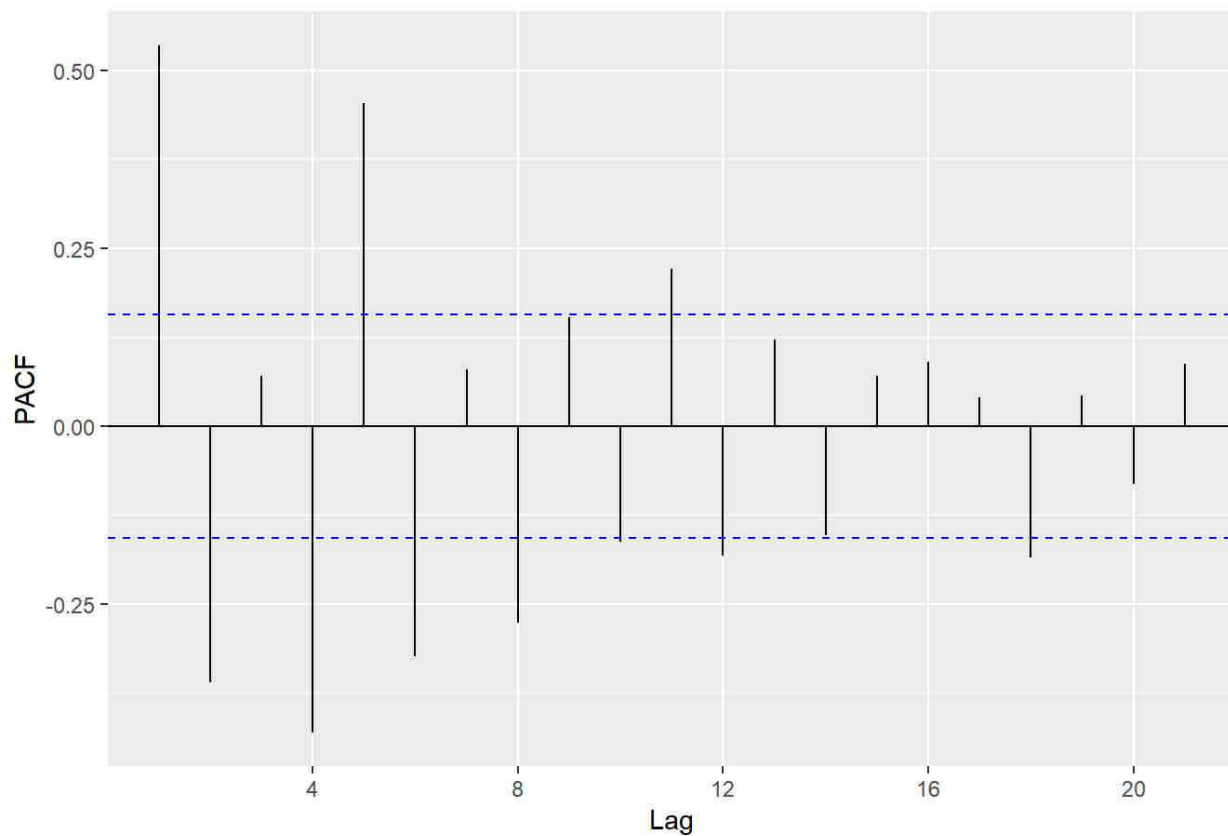
Series: diff(lpib, differences = 2)



```
ggPacf(diff(lpib, differences = 2))
```

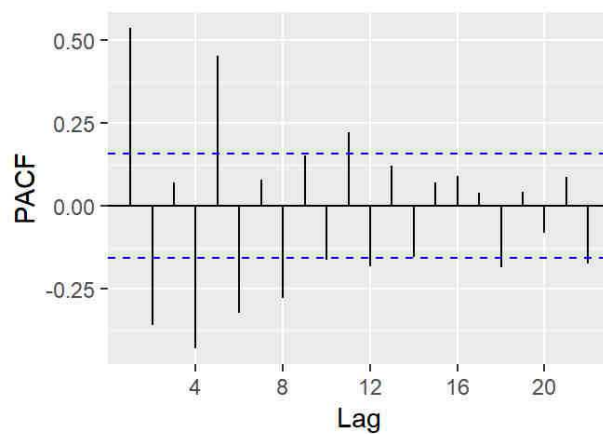
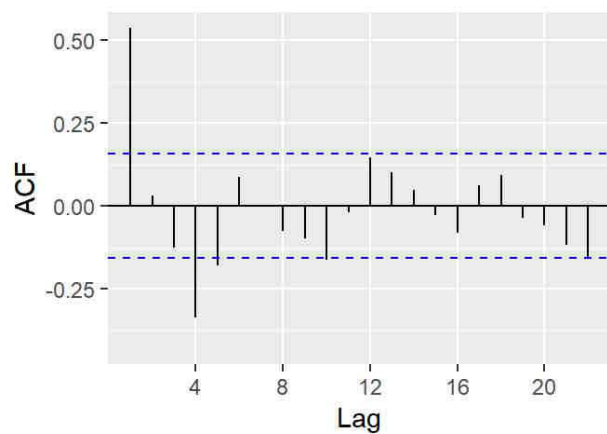
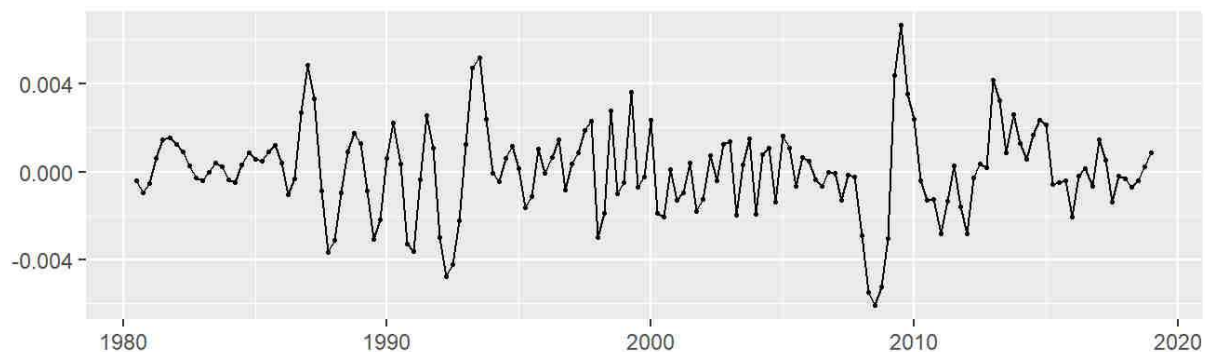


Series: diff(lpib, differences = 2)



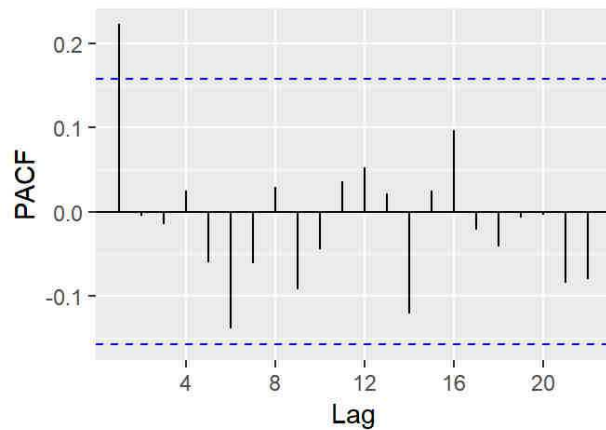
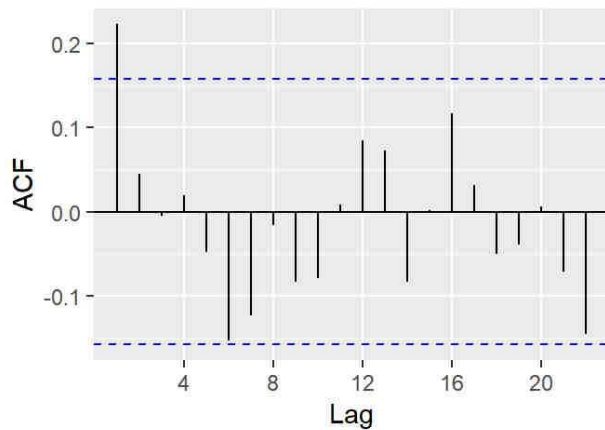
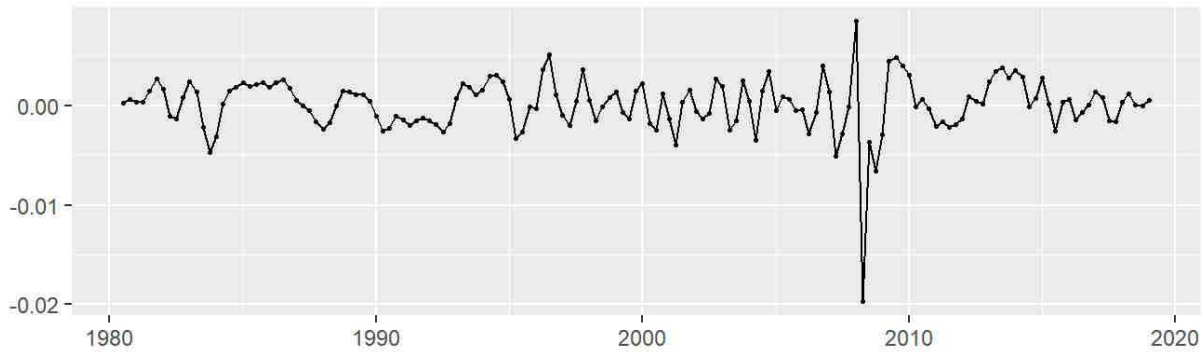
```
ggtsdisplay(diff(lpib, differences = 2))
```

diff(lpib, differences = 2)



```
ggtsdisplay(diff(locu, differences = 2))
```

diff(locu, differences = 2)



# \_\_\_\_\_ #

## 5 Análisis de Cointegración

```
# Para poder seguir supondremos que las dos series son I(2)
# Opción 1: Diferencio las series y compruebo que el modelo siga siendo válido

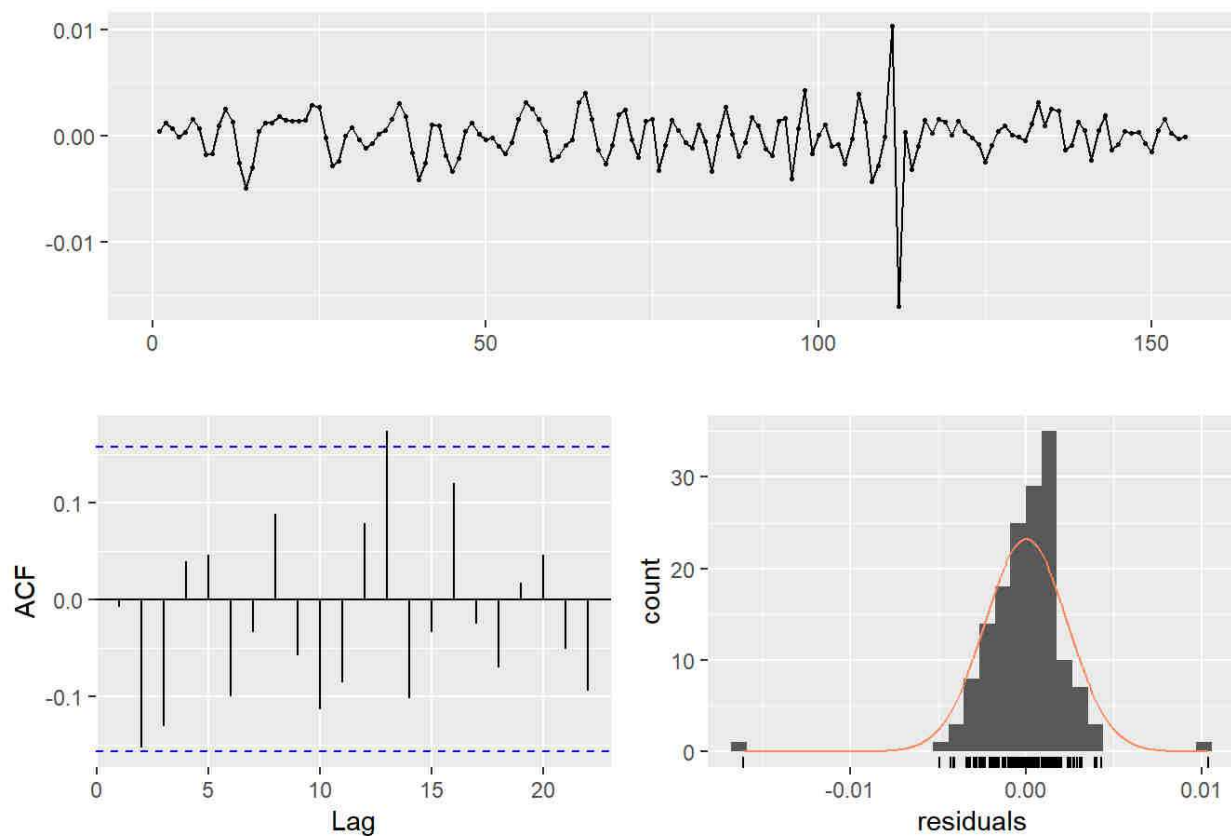
# Construyo las series diferenciadas dy dx (tasas de aceleración - 2 diferencias)
d2locu <- diff(locu, differences=2)
d2lpib <- diff(lpib, differences=2)

# ¿Sigue siendo una relación significativa?
dif.reg <- lm(d2locu~d2lpib)
summary(dif.reg)
```

```
##
## Call:
## lm(formula = d2locu ~ d2lpib)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.0161160 -0.0010167  0.0002156  0.0013631  0.0103617
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  7.094e-05  1.895e-04   0.374   0.709
## d2lpib       6.703e-01  9.396e-02   7.133 3.66e-11 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.002358 on 153 degrees of freedom
## Multiple R-squared:  0.2496, Adjusted R-squared:  0.2447
## F-statistic: 50.88 on 1 and 153 DF,  p-value: 3.66e-11
```

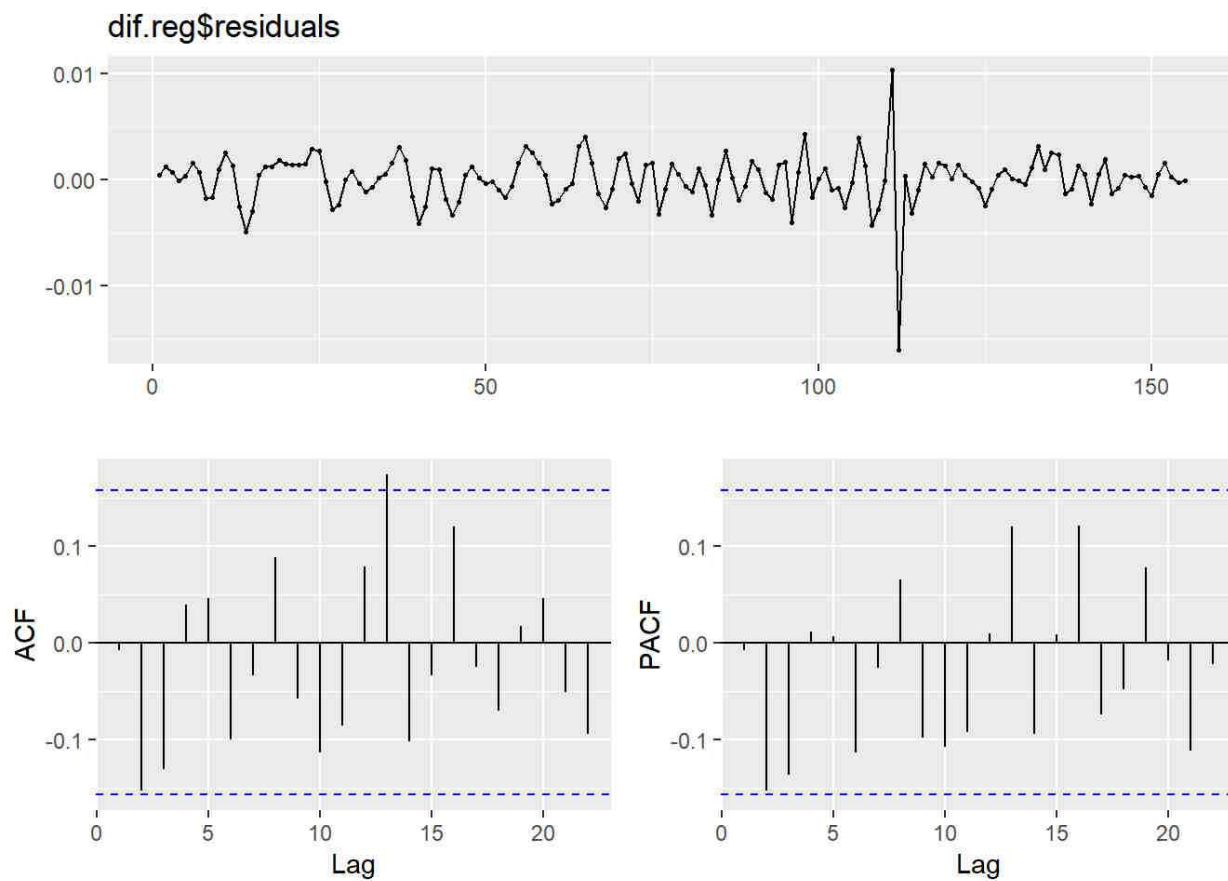
```
# Análisis de los residuos
# Gráfico residuos, acf, histograma
checkresiduals(dif.reg, test="LB")
```

### Residuals



```
##
## Ljung-Box test
##
## data: residuals
## Q* = 12.943, df = 8, p-value = 0.1138
##
## Model df: 2. Total lags used: 10
```

```
# Gráfico residuos, acf, pacf
ggtsdisplay(dif.reg$residuals) # Objeto ggplot
```



```

# ¿Son estacionarios los residuos?

# COMPLETAD VOSOTROS APLICANDO TEST ADF

# _____#
# PIB y ocupados están cointegrados (existe una c.l. de orden inferior: I(1) o I(0))
# ¿Cuál es la relación de causalidad?

# COMPLETAD VOSOTROS

# _____#
# ¿Cómo se estima la relación a L/P de las variables?
# Esto sirve si Y e X son I(1), si son de orden superior (como en este caso), el problema se complica
#
# Estimo  $y = a + bx + u$ 

# Si los residuos son I(0) (existe una c.l. de Y e X que es estacionaria:  $u = y - a - bx$ )
# entonces Y e X están cointegradas.
# INTERPRETACIÓN: Aunque Y e X crezcan en el tiempo, lo hacen de una forma acompañada,
# de modo que el error entre ambas no crece

# En este caso (residuos son I(0)), el parámetro b (beta) es superconsistente

# IDEA: Probar la cointegración entre dos series I(1) es igual que probar la estacionariedad
# de los residuos

# _____#
# Opción 2: Test cointegración Engle-Granger
# Paso 1 (YA hecha arriba): estimación de la estacionariedad de las series: X, Y ~ I(2)
# Paso 2: prueba de cointegración
# Paso 3: método de corrección de errores (ECM)

# PASO 2:
#coint.reg <- lm(locu ~ lpib-1) # Modelo sin constante
#coint.reg <- lm(locu ~ lpib)
summary(coint.reg)

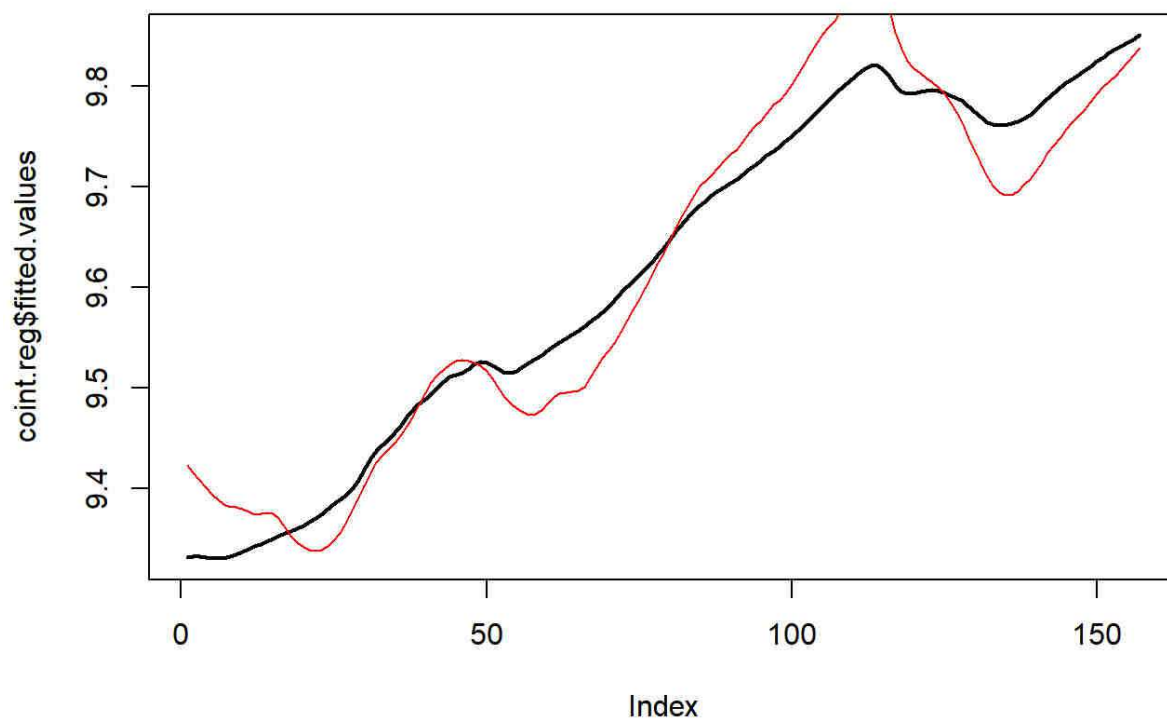
```

```

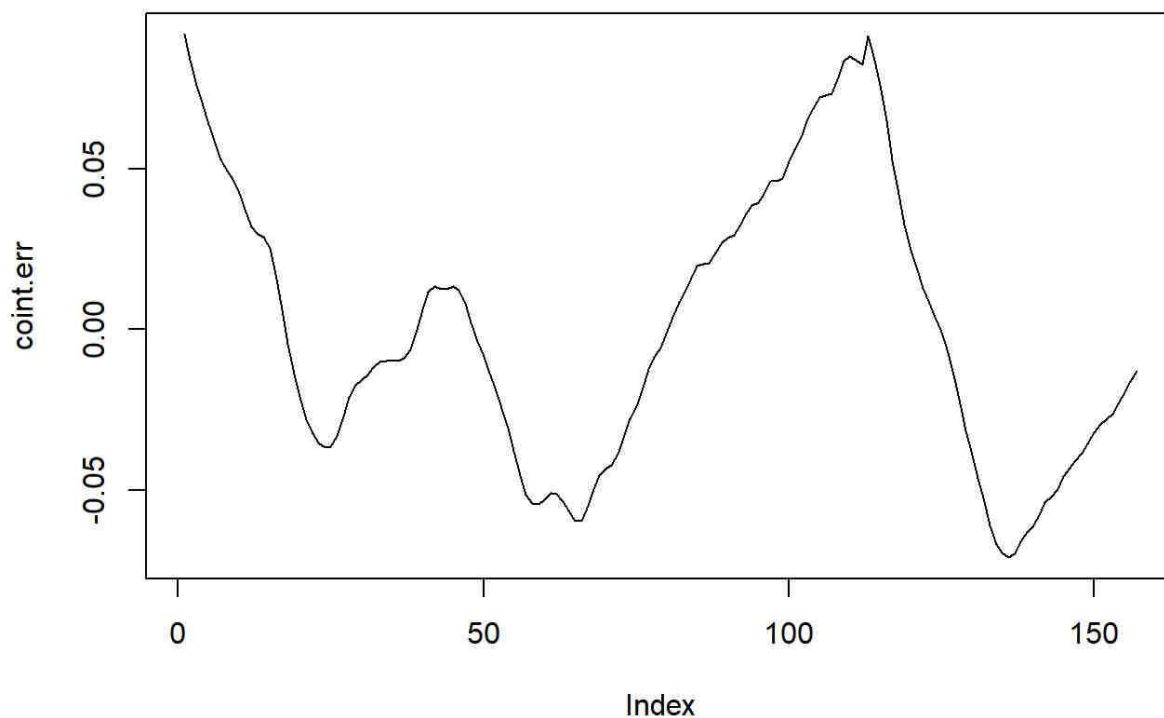
##
## Call:
## lm(formula = locu ~ lpib)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.071068 -0.035477 -0.008451  0.031949  0.091984
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  7.07328    0.05230   135.25  <2e-16 ***
## lpib         0.59107    0.01212    48.75  <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.04383 on 155 degrees of freedom
## Multiple R-squared:  0.9388, Adjusted R-squared:  0.9384
## F-statistic: 2377 on 1 and 155 DF, p-value: < 2.2e-16

```

```
# Representación de los datos reales vs. modelo ajustado
plot(coint.reg$fitted.values, type="l", lwd=2)
lines(coint.reg$model$locu, type="l", col=2)
```



```
# Extraigo el error de equilibrio (residuos)
coint.err <- residuals(coint.reg)
plot(coint.err, type="l")
```



```
# Compruebo si el error es estacionario
# Modo 1: Programa base de R
adf.test(coint.err) # ojo el p-valor no es válido por eso tengo que calcular el p-valor de manera robusta:
```

```
##
## Augmented Dickey-Fuller Test
##
## data: coint.err
## Dickey-Fuller = -2.7951, Lag order = 5, p-value = 0.2449
## alternative hypothesis: stationary
```

```
punitroot(adf.test(coint.err)$statistic, trend="ct")
```

```
## Dickey-Fuller
## 0.1990106
```

```
# Acepto  $H_0$ , hay una raíz unitaria
```

```
adf.test(na.omit(diff(coint.err, differences=1))) # ojo el p-valor no es válido por eso tengo que calcular el p-valor de manera robusta:
```

```
##
## Augmented Dickey-Fuller Test
##
## data: na.omit(diff(coint.err, differences = 1))
## Dickey-Fuller = -2.9678, Lag order = 5, p-value = 0.1729
## alternative hypothesis: stationary
```

```
punitroot(adf.test(na.omit(diff(coint.err, differences=1))))$statistic, trend="ct")
```

```
## Dickey-Fuller
##      0.141383
```

```
# Acepto H0, hay una raíz unitaria
# Eso quiere decir que los residuos son I(2)
# Si los residuos tiene el mismo orden de integración que PIB y OCU
# las variables NO están cointegradas
```

```
# Modo 2: Librería URCA
# ADF: mejor con la librería urca
# En base a phi2 y phi3 no es el modelo elegido
tru <- ur.df(coint.err, type = c("trend"), lags = 12,
             selectlags = c("BIC"))
summary(tru)
```

```
##
## #####
## # Augmented Dickey-Fuller Test Unit Root Test #
## #####
##
## Test regression trend
##
##
## Call:
## lm(formula = z.diff ~ z.lag.1 + 1 + tt + z.diff.lag)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.0142163 -0.0012721  0.0001288  0.0011550  0.0111181
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -1.904e-04  4.377e-04  -0.435  0.66421
## z.lag.1      -1.275e-02  4.625e-03  -2.756  0.00662 **
## tt           1.512e-06  4.632e-06   0.326  0.74467
## z.diff.lag    8.790e-01  4.009e-02  21.927 < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.002306 on 140 degrees of freedom
## Multiple R-squared:  0.7769, Adjusted R-squared:  0.7721
## F-statistic: 162.5 on 3 and 140 DF,  p-value: < 2.2e-16
##
##
## Value of test-statistic is: -2.7564 2.5636 3.8452
##
## Critical values for test statistics:
##      1pct  5pct 10pct
## tau3 -3.99 -3.43 -3.13
## phi2  6.22  4.75  4.07
## phi3  8.43  6.49  5.47
```

```
tru@teststat[1]
```

```
## [1] -2.756439
```



```
punitroot(tru@teststat[1], trend="ct") # p-valor para los residuos
```

```
## [1] 0.2137559
```

```
# No es el modelo seleccionado de acuerdo a phi1
tru <- ur.df(coint.err, type = c("drift"), lags = 12,
            selectlags = c("BIC"))
summary(tru)
```

```
##
## #####
## # Augmented Dickey-Fuller Test Unit Root Test #
## #####
##
## Test regression drift
##
##
## Call:
## lm(formula = z.diff ~ z.lag.1 + 1 + z.diff.lag)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.014182 -0.001258  0.000112  0.001137  0.011159
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -6.236e-05  1.932e-04  -0.323   0.7474
## z.lag.1      -1.274e-02  4.610e-03  -2.763   0.0065 **
## z.diff.lag    8.798e-01  3.988e-02  22.060 <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.002299 on 141 degrees of freedom
## Multiple R-squared:  0.7767, Adjusted R-squared:  0.7735
## F-statistic: 245.2 on 2 and 141 DF,  p-value: < 2.2e-16
##
##
## Value of test-statistic is: -2.7627 3.8163
##
## Critical values for test statistics:
##      1pct  5pct 10pct
## tau2 -3.46 -2.88 -2.57
## phi1  6.52  4.63  3.81
```

```
tru@teststat[1]
```

```
## [1] -2.762677
```

```
punitroot(tru@teststat[1], trend="c") # Aquí depende del nivel de significación que escoja
```

```
## [1] 0.06375145
```

```
# Modelo seleccionado
tru <- ur.df(diff(coint.err, differences=1), type = c("none"), lags = 12,
             selectlags = c("BIC"))
summary(tru)
```

```
##
## #####
## # Augmented Dickey-Fuller Test Unit Root Test #
## #####
##
## Test regression none
##
##
## Call:
## lm(formula = z.diff ~ z.lag.1 - 1 + z.diff.lag)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.0159794 -0.0009952  0.0000087  0.0010886  0.0100319
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## z.lag.1      -0.13185    0.04209  -3.132  0.00211 **
## z.diff.lag   0.06606    0.08405   0.786  0.43321
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.002353 on 141 degrees of freedom
## Multiple R-squared:  0.06506,    Adjusted R-squared:  0.0518
## F-statistic: 4.906 on 2 and 141 DF,  p-value: 0.008713
##
##
## Value of test-statistic is: -3.1324
##
## Critical values for test statistics:
##      1pct  5pct 10pct
## tau1 -2.58 -1.95 -1.62
```

```
punitroot(tru@teststat[1], trend="nc")
```

```
## [1] 0.001694471
```

```
# Rechazo H0 (p-valor 0.001499455). Eso quiere decir que NO HAY
# una raíz unitaria
# Eso quiere decir que los residuos de mi modelo son estacionarios, I(0)

# Como el orden de integración de los residuos I(0) es inferior a 2,
# las series de PIB y OCU están cointegradas
# _____#
```

## 6 Modelo de Corrección del error

```
# PASO 3: ECM
# ¿Cuál es la conclusión?
# Contruyo la serie de residuos retardada

length(lpib)
```

```
## [1] 157
```

```
coint.err.lag <- coint.err[1:(length(coint.err)-1)]
length(coint.err.lag)
```

```
## [1] 156
```

```
# Construyo las series diferenciadas dy dx
dy <- diff(locu, differences=1)
dx <- diff(lpib, differences=1)
length(dy)
```

```
## [1] 156
```

```
# Construyo la serie dyt-1, dyt-1
dy.lag <- dy[1:(length(dy)-1)]
dx.lag <- dx[1:(length(dx)-1)]
length(dy.lag)
```

```
## [1] 155
```

```
# Ajusto el tamaño de las series para que todas tenga el mismo tamaño
dy <- dy[2:length(dy)]
dx <- dx[2:length(dx)]
coint.err.lag <- coint.err.lag[1:length(coint.err.lag)-1]
head(dy)
```

```
## [1] -0.007554291 -0.006938719 -0.006563702 -0.006200731 -0.004775826
## [6] -0.002144933
```

```
tail(dy)
```

```
## [1] 0.005207547 0.005566434 0.006717152 0.006733130 0.006651717 0.007220793
```

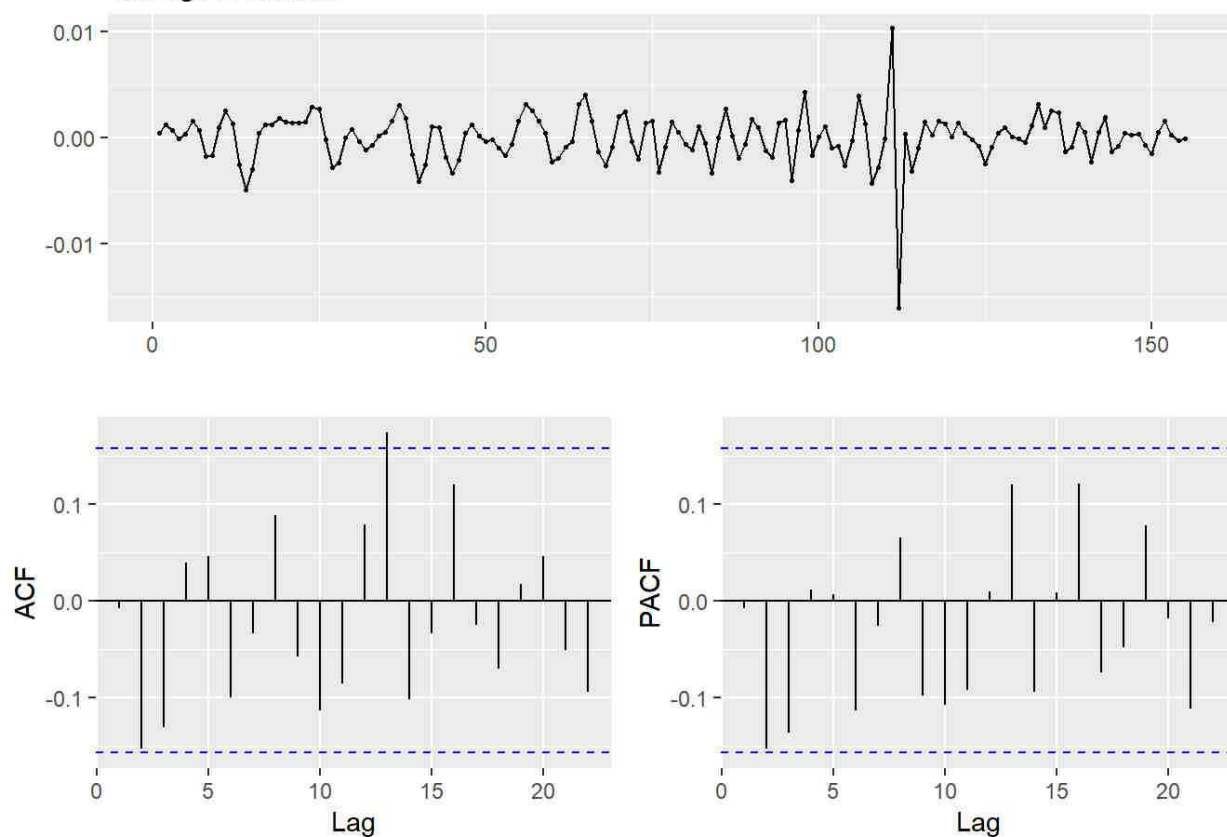
```
# Estimo el Modelo de corrección de Error (ECM):
# Solo con un retardo para las diferencias de Y y de X
ecm <- lm(dy ~ coint.err.lag + dy.lag + dx.lag)
summary(ecm)
```

```
##
## Call:
## lm(formula = dy ~ coint.err.lag + dy.lag + dx.lag)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.0156679 -0.0013032  0.0001304  0.0014087  0.0093783
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  -0.0007718   0.0003848   -2.006  0.046675 *
## coint.err.lag -0.0172620   0.0046608   -3.704  0.000298 ***
## dy.lag        0.7228378   0.0633863   11.404 < 2e-16 ***
## dx.lag        0.2843994   0.0839880    3.386  0.000904 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.0025 on 151 degrees of freedom
## Multiple R-squared:  0.898, Adjusted R-squared:  0.896
## F-statistic: 443.1 on 3 and 151 DF, p-value: < 2.2e-16
```

```
# Compruebo que el signo de coint.err.lag es el correcto (negativo)
# El signo de dx.lag me da la relación a C/P entre y e x

# Diagnósis de los residuos
# Tienen que ser estacionarios
ggtsdisplay(dif.reg$residuals) # Objeto ggplot
```

dif.reg\$residuals



```
# ¿Son estacionarios los residuos?  
# Completad vosotros con test ADF  
  
# _____ #  
# La librería "dynlm" nos simplifica la vida  
# _____ #  
class(ocu)
```

```
## [1] "ts"
```

```
class(lpib)
```

```
## [1] "ts"
```

```
class(coint.err)
```

```
## [1] "numeric"
```

```
coint.err <- ts(coint.err,start=1980,frequency=4)  
class(coint.err)
```

```
## [1] "ts"
```

```
# Creo primero la serie de diferencias  
diff.ocu <- diff(ocu,differences = 1)  
diff.pib <- diff(lpib,differences = 1)  
  
# OJO: En el comando de lags L(), si solo hay un retardo no hace falta especificarlo  
dynecm <- dynlm(diff.ocu ~ L(coint.err) + L(diff.ocu,1) + L(diff.pib,1))  
summary(dynecm)
```

```
##
## Time series regression with "ts" data:
## Start = 1980(3), End = 2019(1)
##
## Call:
## dynlm(formula = diff.ocu ~ L(coint.err) + L(diff.ocu, 1) + L(diff.pib,
##    1))
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.0156679 -0.0013032  0.0001304  0.0014087  0.0093783
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)   -0.0007718  0.0003848  -2.006 0.046675 *
## L(coint.err)   -0.0172620  0.0046608  -3.704 0.000298 ***
## L(diff.ocu, 1)  0.7400997  0.0630257  11.743 < 2e-16 ***
## L(diff.pib, 1)  0.2741964  0.0839079   3.268 0.001342 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.0025 on 151 degrees of freedom
## Multiple R-squared:  0.898, Adjusted R-squared:  0.896
## F-statistic: 443.1 on 3 and 151 DF, p-value: < 2.2e-16
```

```
# En un solo paso, sin construir previamente la serie de diferencias
dynecm1 <- dynlm(d(ocu) ~ L(coint.err) + L(d(ocu),1) + L(d(lpib),1))
summary(dynecm1)
```

```
##
## Time series regression with "ts" data:
## Start = 1980(3), End = 2019(1)
##
## Call:
## dynlm(formula = d(ocu) ~ L(coint.err) + L(d(ocu), 1) + L(d(lpib),
##    1))
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.0156679 -0.0013032  0.0001304  0.0014087  0.0093783
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)   -0.0007718  0.0003848  -2.006 0.046675 *
## L(coint.err)   -0.0172620  0.0046608  -3.704 0.000298 ***
## L(d(ocu), 1)   0.7400997  0.0630257  11.743 < 2e-16 ***
## L(d(lpib), 1)  0.2741964  0.0839079   3.268 0.001342 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.0025 on 151 degrees of freedom
## Multiple R-squared:  0.898, Adjusted R-squared:  0.896
## F-statistic: 443.1 on 3 and 151 DF, p-value: < 2.2e-16
```

```
stargazer(ecm,dynecm,dynecm1, type="text")
```

```
##
## =====
##                               Dependent variable:
##                               -----
##                               dy      diff.ocu  d(locu)
##                               OLS      dynamic  dynamic
##                               linear    linear
##                               (1)      (2)      (3)
## -----
## coint.err.lag      -0.017***
##                   (0.005)
##
## dy.lag             0.723***
##                   (0.063)
##
## dx.lag             0.284***
##                   (0.084)
##
## L(coint.err)              -0.017***  -0.017***
##                          (0.005)  (0.005)
##
## L(diff.ocu, 1)           0.740***
##                          (0.063)
##
## L(diff.pib, 1)          0.274***
##                          (0.084)
##
## L(d(locu), 1)                  0.740***
##                          (0.063)
##
## L(d(lpib), 1)              0.274***
##                          (0.084)
##
## Constant              -0.001**  -0.001**  -0.001**
##                      (0.0004)  (0.0004)  (0.0004)
##
## -----
## Observations           155         155         155
## R2                     0.898         0.898         0.898
## Adjusted R2            0.896         0.896         0.896
## Residual Std. Error (df = 151)  0.002         0.002         0.002
## F Statistic (df = 3; 151)  443.056***  443.056***  443.056***
## =====
## Note:                  *p<0.1; **p<0.05; ***p<0.01
```

```
# ¿Por qué me da distinto "ecm"? Por favor, revisad que subíndices sean correctos
dim(model.frame(ecm))
```

```
## [1] 155  4
```

```
dim(model.frame(dynecm))
```

```
## [1] 155  4
```

```
dim(model.frame(dynecm1))
```

```
## [1] 155    4
```

```
# COMPLETAD VOSOTROS
```

```
# Seleccionar el modelo ECM final, ¿cuántos retardos incluyo de diferencias de ocupados y pib?
```

## 7 Posibles ampliaciones

```
# _____#  
# FORECASTING  
# Desgraciadamente, NO EXISTE el comando predict() asociado a la librería dynlm()  
# https://stats.stackexchange.com/questions/6758/1-step-ahead-predictions-with-dynlm-r-package  
# _____#  
# CAMBIO ESTRUCTURAL Y TESTS DE RAÍCES UNITARIAS  
# Malas noticias: Todos los tests descritos son ineficientes en el caso de cambio estructural.  
# Hay solución pero está fuera del alcance de este curso (avanzado)
```





Proyecto de innovación  
“Innova Docencia” de la UCM  
Convocatoria 2018-2019  
Proyecto número 275

**R-adaptación de la asignatura de métodos econométricos en economía y finanzas del  
Grado de Estadística Aplicada**

Facultad de Estudios Estadísticos  
Departamento de Economía Aplicada, Pública y Política

**Título**

***Práctica 11: Autocorrelación en La Ley de Okun***

**Correspondiente al**

**Capítulo 5 MODELOS CON DATOS DE SERIES TEMPORALES II: Autocorrelación,  
indicadores adelantados y causalidad en sentido de Granger**

- Modelos de regresión con **series temporales estacionarias**: autocorrelación
- Contrastes de autocorrelación
- El estimador de Mínimos Cuadrados Generalizados (MCG)  
***Práctica 11: Autocorrelación en La Ley de Okun***
- Indicadores adelantados: Causalidad en sentido de Granger y  
correlación contemporánea  
***Práctica 12: El Índice de Confianza del Consumidor como indicador  
adelantado del PIB***

# Modelos de regresión con series temporales estacionarias y Autocorrelación en los residuos

## Asignatura Métodos Econométricos en Economía y Finanzas

Profesores  
Loren Escot, Alicia Pérez-Alonso y Julio E. Sandubete

Dpto Economía Aplicada, Pública y Política  
Facultad Estudios Estadísticos - Universidad Complutense de Madrid

Curso académico 2018/2019

### Abstract

En esta práctica se muestra cómo detectar problemas de autocorrelación en el término de error de un modelo de regresión con series temporales estacionarias. A modo de ejemplo retomamos la regresión de la ley de Ley de Okun con las tasas de crecimiento de los ocupados y del PIB.  
Una vez detectado un problema de autocorrelación, mostraremos las diferentes alternativas para tratar con el problema de la autocorrelación en el término de error

- 1 Librerías necesarias
- 2 Carga de datos
- 3 Paso 1: Modelo de regresión lineal (bajo hipótesis MRLC)
- 4 Paso 2: Diagnóstico residuos para ver si se cumplen hipótesis MRLC (AUTOCORRELACIÓN)
- 5 Paso 3: Contrastes de Autocorrelación en los residuos
  - 5.1 Test de Durbin-Watson
  - 5.2 Test de Breusch-Godfrey
  - 5.3 Test de Ljung-Box
  - 5.4 CORRELOGRAMA
    - 5.4.1 Correlograma con la Librería "forecast" (varias opciones)
- 6 Paso 4: Estimación del modelo con autocorrelación en el error
  - 6.1 (1): Estimador MCO pero con un estimador consistente (Newey-West) de la matriz var-cov de los coeficientes MCO
  - 6.2 (2) Mínimos Cuadrados Generalizados Factible - errores AR(1)
  - 6.3 (3) Estimo modelo lineal con un proceso AR(1) en el término de error
  - 6.4 (4) Estimo modelo lineal con un proceso AR(2) en el término de error
  - 6.5 (5) Estimo modelo lineal con un proceso auto.arima() en el término de error

```
#=====#
#   File: t3autocorr.R
#   This script was written for R.3.4.3
#   First version: 12abr2018
#   Last version: 01abr2019
#   Last revisor:
#_____#
#   Description: Autocorrelación
#   TC_ocu(t) = B1 + B2* TC_pib(t) + u(t);   "Ley de Okun"
#   u(t) ~ ARIMA(p,d,q)
#
#   Packages: stargazer, ggplot2, lmtest, sandwich, forecast, orcutt, prais
#   Input: okundat.Rdata  [DATOS EMPLEO EQUIVALENTE]
#   Output:
#
#   References:
#   http://ggplot2.tidyverse.org/reference/ggplot.html
#   http://www.sthda.com/english/wiki/print.php?id=188
#   https://stackoverflow.com/questions/29554796/meaning-of-band-width-in-ggplot-geom-smooth-lm
#
#   https://robjhyndman.com/hyndsight/arimax/
#   https://robjhyndman.com/hyndsight/arimaconstants/
#   https://www.otexts.org/fpp/8/7
#   #https://github.com/cran/prais/blob/master/R/prais.winsten.R
#=====#
```

# 1 Librerías necesarias

```
# Librerías necesarias ----
#dev.off()                # Cierra gráficos que podamos tener abiertos
rm(list=ls())             # Elimina todos los objetos que tengamos en la memoria
```

Para instalar los paquetes si no se encuentran  
(ya no es necesario en las nuevas versiones de Rstudio)

```
#if(!require(stargazer)){install.packages("stargazer")}
#if(!require(ggplot2)){install.packages("ggplot2")}
#if(!require(lmtest)){install.packages("lmtest")}
#if(!require(sandwich)){install.packages("sandwich")}
#if(!require(forecast)){install.packages("forecast")}
#if(!require(orcutt)){install.packages("orcutt")}
#if(!require(prais)){install.packages("prais")}

suppressMessages(library(stargazer))
suppressMessages(library(ggplot2))
suppressMessages(library(lmtest))                # Contrastes autocorrelación
suppressMessages(library(sandwich))              # Estimador consistente matriz var-cov de los esti
  madores MCO
suppressMessages(library(forecast))              # Arima(); Acf(); Pacf() => mejoran comandos base
  de R + auto.arima()
suppressMessages(library(orcutt))                # Cochrane-Orcutt (mínimos cuadrados generalizados
  factibles)
suppressMessages(library(prais))                 # Prais-Winsten (mínimos cuadrados generalizados f
  actibles)

#_____#
```

## 2 Carga de datos

```
# Datos (Ley de Okun) ----  
load("okundat.Rdata")  
head(okun)
```

```
##   time.pib.ts.  pib.ts  ocu.ts   T1pib.ts   T4pib.ts  T1ocu.ts  T4ocu.ts  
## 1    1970.00 32.18634    NA        NA        NA        NA        NA  
## 2    1970.25 32.23748    NA 0.001588640    NA        NA        NA  
## 3    1970.50 32.40224    NA 0.005110833    NA        NA        NA  
## 4    1970.75 32.65790    NA 0.007890276    NA        NA        NA  
## 5    1971.00 32.91293    NA 0.007809178 0.02257438    NA        NA  
## 6    1971.25 33.55493    NA 0.019505927 0.04086708    NA        NA
```

```
tail(okun)
```

```
##   time.pib.ts.  pib.ts  ocu.ts   T1pib.ts   T4pib.ts   T1ocu.ts  
## 191    2017.50 105.8290 18027.35 0.006392359 0.02924071 0.006761089  
## 192    2017.75 106.5982 18118.42 0.007268329 0.03089747 0.005051657  
## 193    2018.00 107.1900 18208.20 0.005551688 0.02824484 0.004955123  
## 194    2018.25 107.7914 18347.10 0.005610598 0.02505401 0.007628486  
## 195    2018.50 108.3850 18477.13 0.005506933 0.02415217 0.007087278  
## 196    2018.75 109.1395 18581.26 0.006961295 0.02383999 0.005635832  
##           T4ocu.ts  
## 191 0.02914501  
## 192 0.02869737  
## 193 0.02564363  
## 194 0.02461784  
## 195 0.02494981  
## 196 0.02554555
```

```
T4pib.ts <- window(okun[,5], start=1981, end=2018.50)  
T4ocu.ts <- window(okun[,7], start=1981, end=2018.50)  
datokun <- ts.intersect(T4pib.ts,T4ocu.ts,dframe = TRUE)
```

Podemos exportar los datos a otros formatos: .txt, .csv se quiereos utilizarlos en otro software

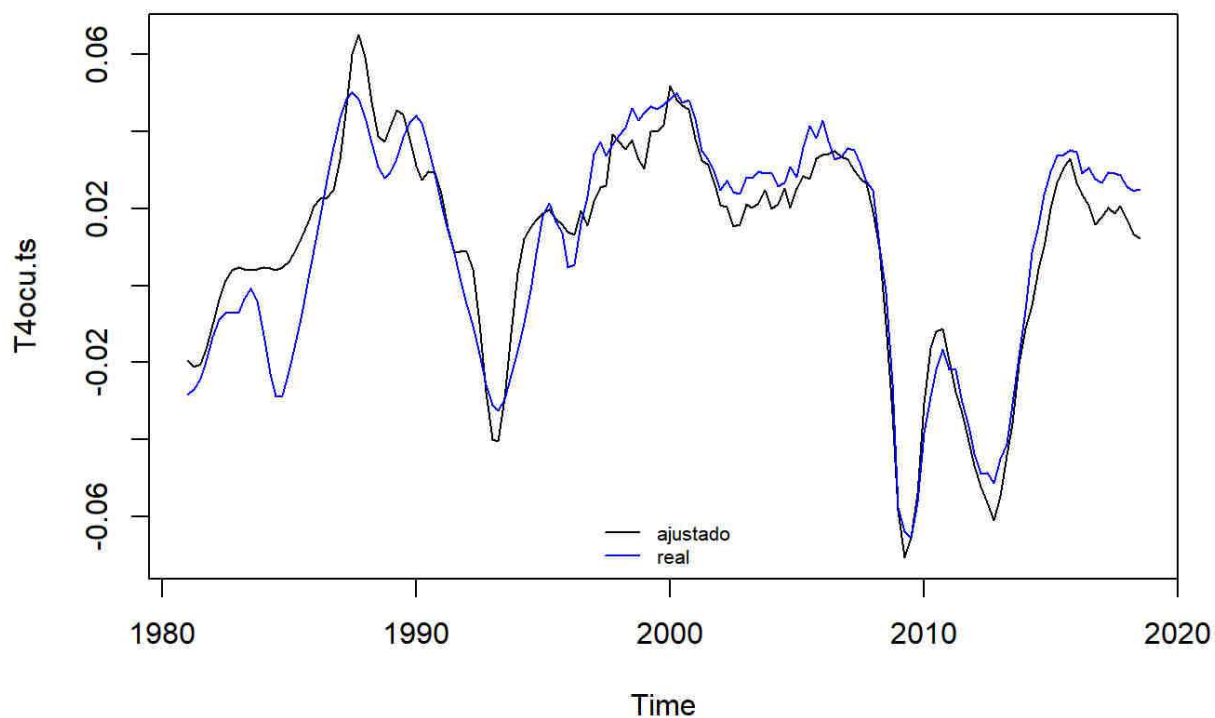
```
# write.table(datokun, file="okundat.txt", row.names = FALSE, sep=",")  
#_____#
```

## 3 Paso 1: Modelo de regresión lineal (bajo hipótesis MRLC)

```
# Paso 1: Modelo de regresión lineal (bajo hipótesis MRLC)----  
modelo <- T4ocu.ts ~ T4pib.ts  
modelo.lm <- lm(modelo, data=datokun, na.action=NULL) #ojo, na.action=NULL para las series  
salida.lm <- summary(modelo.lm)  
print(salida.lm)
```

```
##
## Call:
## lm(formula = modelo, data = datokun, na.action = NULL)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.033238 -0.005828  0.002526  0.007613  0.014675
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -0.017713   0.001169  -15.15  <2e-16 ***
## T4pib.ts     1.243486   0.036155   34.39  <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.009997 on 149 degrees of freedom
## Multiple R-squared:  0.8881, Adjusted R-squared:  0.8874
## F-statistic: 1183 on 1 and 149 DF, p-value: < 2.2e-16
```

```
# Representación gráfica de la estimación
plot(fitted(modelo.lm), ylab="T4ocu.ts")
lines(T4ocu.ts,col="blue")
legend("bottom",legend = c("ajustado", "real"), lty=c(1), cex=0.65, lwd=c(1,1),
      bty="n",
      col=c("black","blue"))
```



# \_\_\_\_\_ #

## 4 Paso 2: Diagnósis residuos para ver si se cumplen hipótesis MRLC (AUTOCORRELACIÓN)

```
# Paso 2: Diagnósis residuos para ver si se cumplen hipótesis MRLC----
```

```
# Obtención de los residuos
```

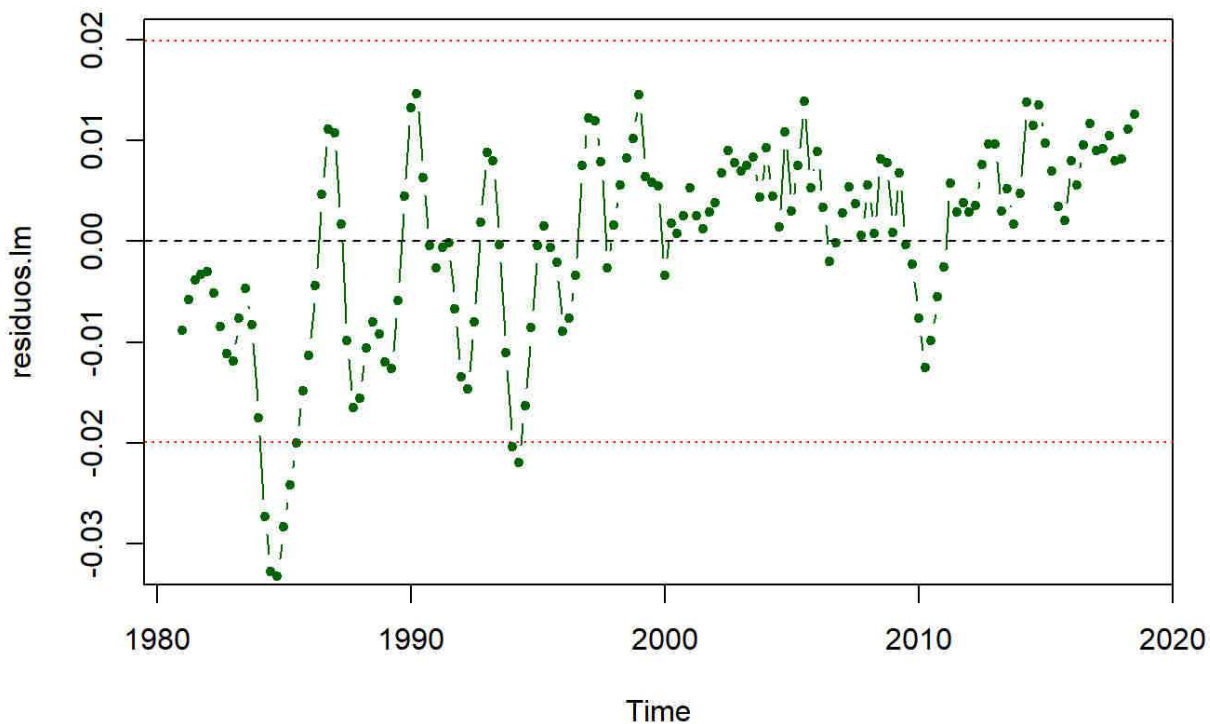
```
residuos.lm <- residuals(modelo.lm)  
class(residuos.lm) # Objeto serie temporal
```

```
## [1] "ts"
```

```
# ¿Son los residuos un Ruido Blanco Gaussiano?  
summary(residuos.lm)
```

```
##      Min.    1st Qu.    Median      Mean   3rd Qu.      Max.     
## -0.033238 -0.005828  0.002526  0.000000  0.007613  0.014675
```

```
plot(residuos.lm, type="b", col="darkgreen", pch=20, ylim=c(-0.032,0.02))  
abline(h=0, lty=2, col="grey2" )  
abline(h=2*sd(residuos.lm), lty=3, col="red" )  
abline(h=-2*sd(residuos.lm), lty=3, col="red" )
```



Trabajamos con series temporales, por lo que sospechamos de la existencia de autocorrelación en el término de error.

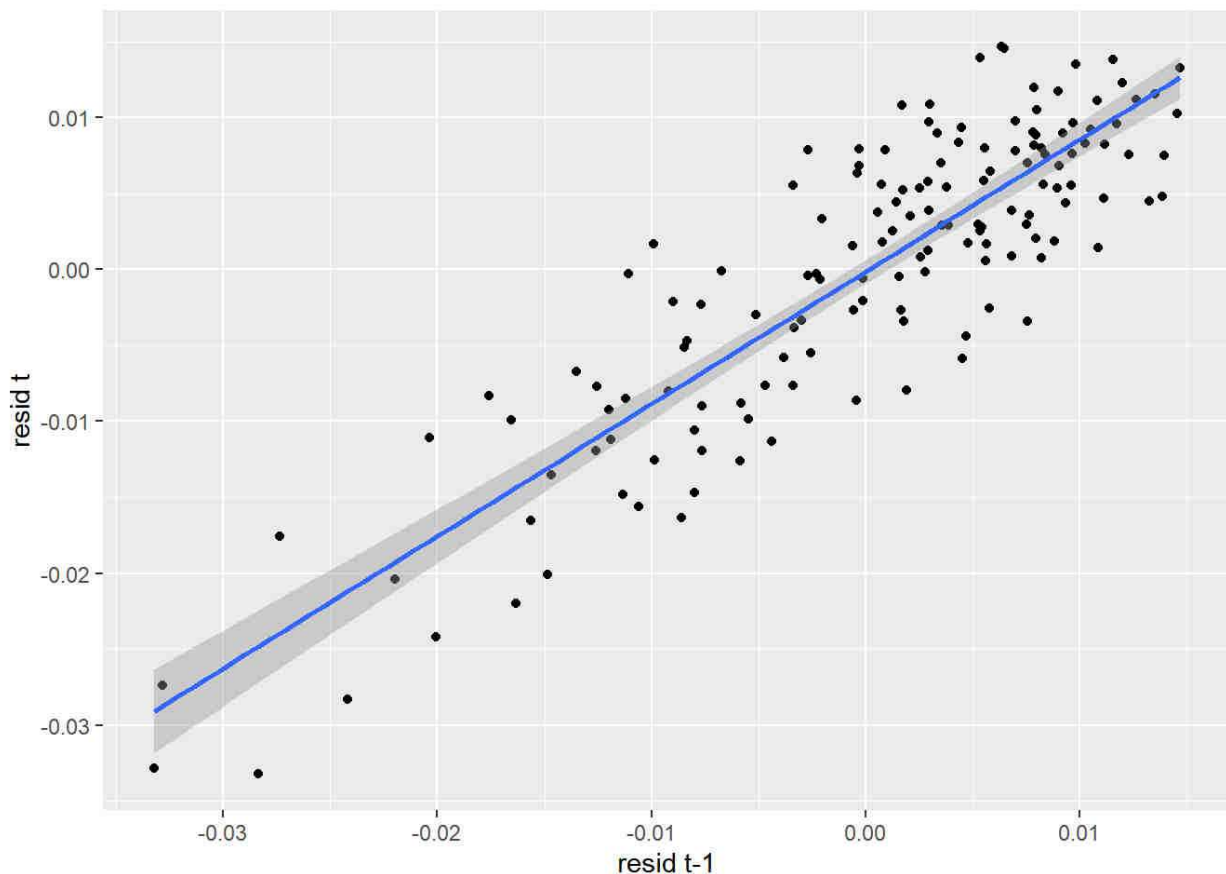
Por tanto en la diagnosis de los residuos conviene hacer algún gráfico adicional y pasarle a los residuos los test formales de autocorrelación ## Gráfico de los residuos(t) vs residuos(t-1) Comparamos los residuos con los residuos retardados un periodo (pierdo una observación)

```
residuos_plot <- ts.intersect("resid_t"=residuos.lm, "resid_t1"=lag(residuos.lm, k = 1), dframe=TRUE
)
class(residuos_plot) # Objeto data.frame
```

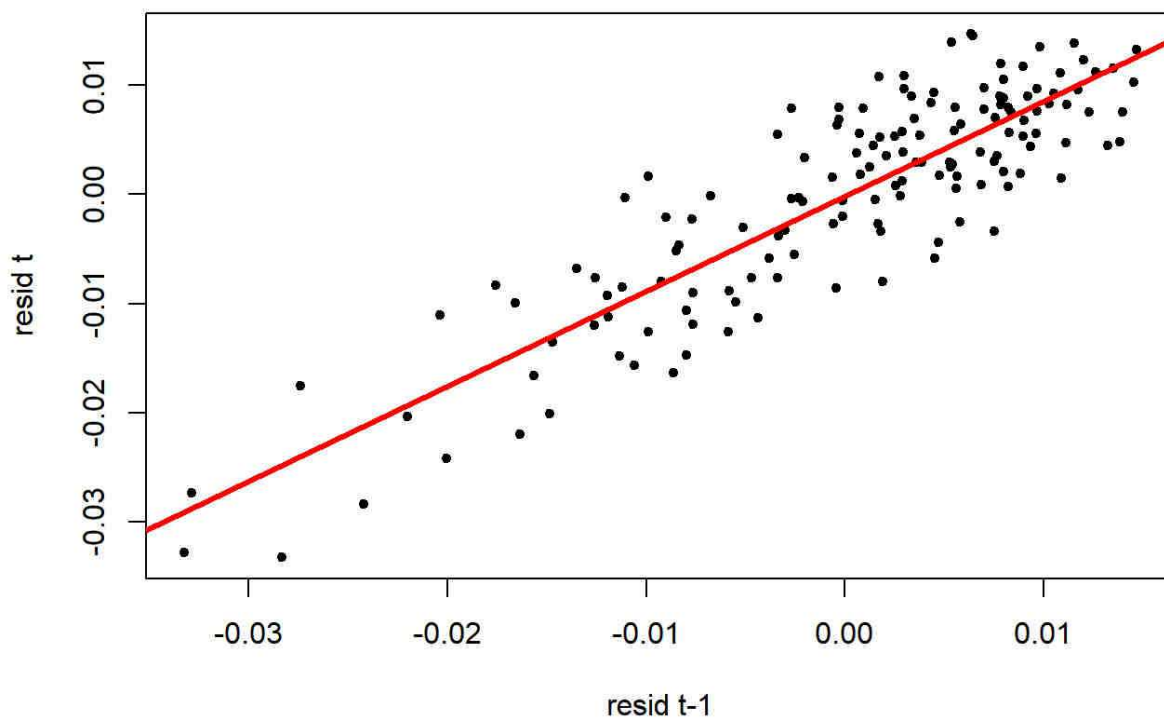
```
## [1] "data.frame"
```

```
# Regresión: residuos(t)/residuos(t-1)
# Alternativa 1 (objeto ggplot)
ggplot(data=residuos_plot) +
  geom_point(aes(x=resid_t1, y=resid_t)) +
  xlab("resid t-1") +
  ylab("resid t") +
  geom_smooth(aes(x=resid_t1, y=resid_t), method = "lm")
```

```
## Don't know how to automatically pick scale for object of type ts. Defaulting to continuous.
## Don't know how to automatically pick scale for object of type ts. Defaulting to continuous.
```



```
# Alternativa 2
plot(as.numeric(residuos_plot[,1])~as.numeric(residuos_plot[,2]),
     xlab="resid t-1", ylab="resid t", type="p", pch=20)
abline(lm(as.numeric(residuos_plot[,1])~as.numeric(residuos_plot[,2]), data=residuos_plot), col=2, l
       wd=3)
```



```
# _____ #
```

## 5 Paso 3: Contrastes de Autocorrelación en los residuos

```
# Paso 3: Contrastes de Autocorrelación en los residuos ----
```

La mayoría de estos test de autocorrelación contrastan la hipótesis nula **H0: Ausencia de Autocorrelación**

### 5.1 Test de Durbin-Watson

Test de Durbin-Watson

**H0: Ausencia de autocorrelación de orden 1**

```
# Durbin-Watson Test
# H0: ausencia de autocorrelación de orden 1
lmtest::dwtest(modelo.lm)
```

```
##
## Durbin-Watson test
##
## data: modelo.lm
## DW = 0.25147, p-value < 2.2e-16
## alternative hypothesis: true autocorrelation is greater than 0
```

```
#lmtest::dwtest(modelo.lm, alternative="less")
```



## 5.2 Test de Breusch-Godfrey

Breusch-Godfrey Test

**H0: ausencia de autocorrelación de orden p**

```
# Breusch-Godfrey Test
# H0: ausencia de autocorrelación de orden p
lmtest::bgtest(modelo.lm, order = 1)
```

```
##
## Breusch-Godfrey test for serial correlation of order up to 1
##
## data: modelo.lm
## LM test = 114.56, df = 1, p-value < 2.2e-16
```

```
lmtest::bgtest(modelo.lm, order = 2)
```

```
##
## Breusch-Godfrey test for serial correlation of order up to 2
##
## data: modelo.lm
## LM test = 119.26, df = 2, p-value < 2.2e-16
```

Ojo!!!!, este test está diseñado para trabajar sólo con residuos de modelos lm. Falla si se intenta aplicar sobre residuos de otros modelos de regresión

## 5.3 Test de Ljung-Box

Ljung-Box Test **H0: ausencia de autocorrelación de orden p**

otras formas de interpretar la H0

**H0: Las autocorrelaciones hasta orden p son iguales a cero**

**H0: Los datos son independientes (hasta orden p)**

```
# vamos a hacer un bucle para estimar 5 hipótesis
for (i in 1:5){
  print(Box.test(residuos.lm, lag = i, type = c("Ljung-Box")))
}
```

```
##
## Box-Ljung test
##
## data:  residuos.lm
## X-squared = 115.59, df = 1, p-value < 2.2e-16
##
##
## Box-Ljung test
##
## data:  residuos.lm
## X-squared = 183.95, df = 2, p-value < 2.2e-16
##
##
## Box-Ljung test
##
## data:  residuos.lm
## X-squared = 216.87, df = 3, p-value < 2.2e-16
##
##
## Box-Ljung test
##
## data:  residuos.lm
## X-squared = 231.93, df = 4, p-value < 2.2e-16
##
##
## Box-Ljung test
##
## data:  residuos.lm
## X-squared = 246.09, df = 5, p-value < 2.2e-16
```

```
# Librería forecast
# Para objetos "lm", valor predeterminado de test=BG; otherwise test=LB
checkresiduals(modelo.lm, plot=FALSE)
```

```
##
## Breusch-Godfrey test for serial correlation of order up to 8
##
## data:  Residuals
## LM test = 128.03, df = 8, p-value < 2.2e-16
```

```
checkresiduals(modelo.lm, test="LB", plot=FALSE)
```

```
##
## Ljung-Box test
##
## data:  Residuals
## Q* = 288.57, df = 6, p-value < 2.2e-16
##
## Model df: 2.    Total lags used: 8
```

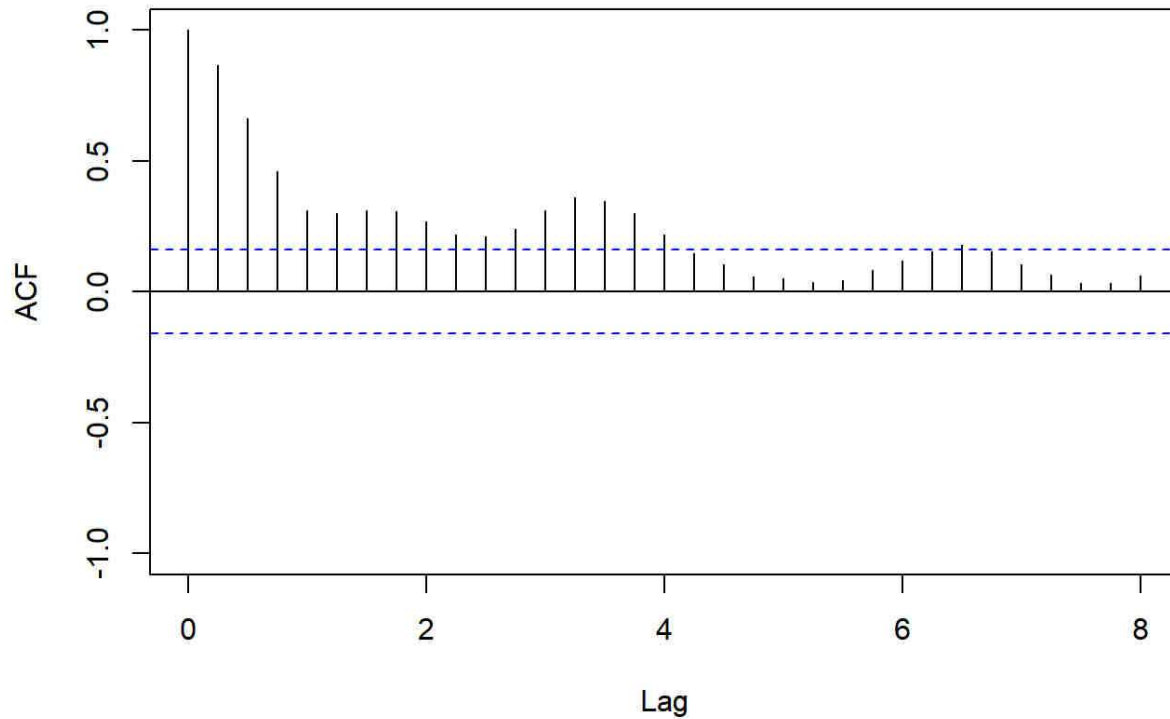
```
# _____ #
```

## 5.4 CORRELOGRAMA

Hay varios paquetes y funciones que permiten calcular el correlograma: función de autocorrelación simple (ACF) y Función de autocorrelación parcial (PACF) ### Correlograma en Base de R

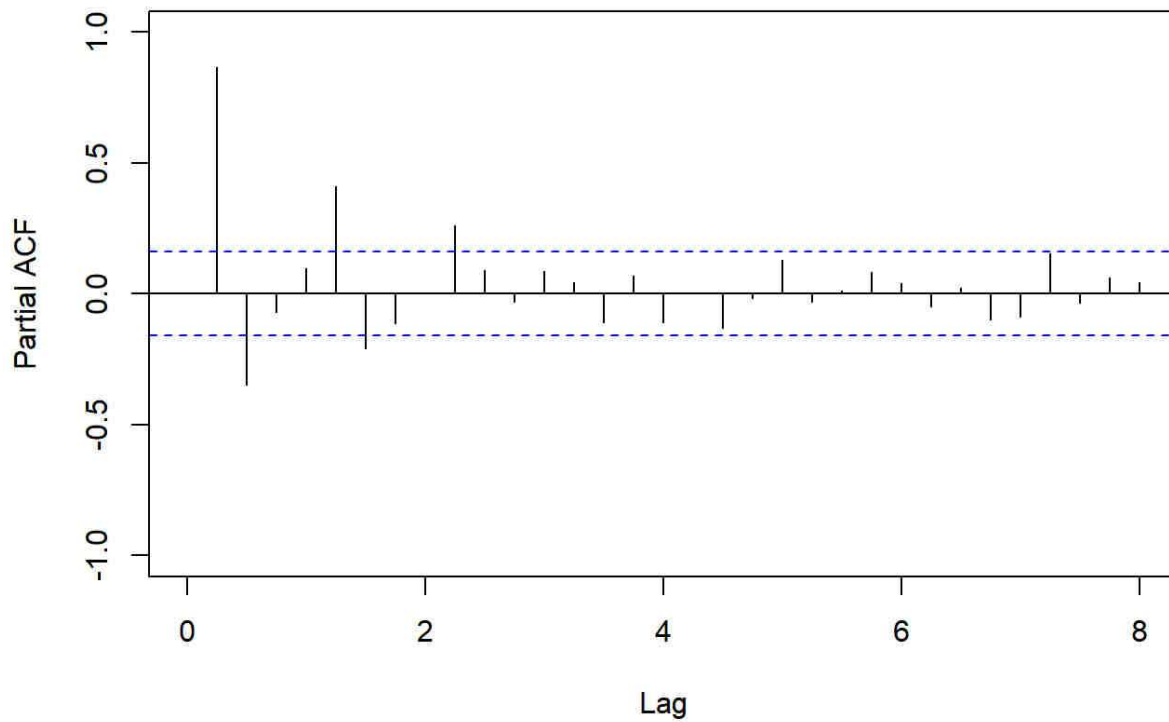
```
acf(residuos.lm, lag.max = 32, xlim=c(0,32/4), ylim=c(-1,1))
```

### Series residuos.lm



```
pacf(residuos.lm, lag.max = 32, xlim=c(0,32/4), ylim=c(-1,1))
```

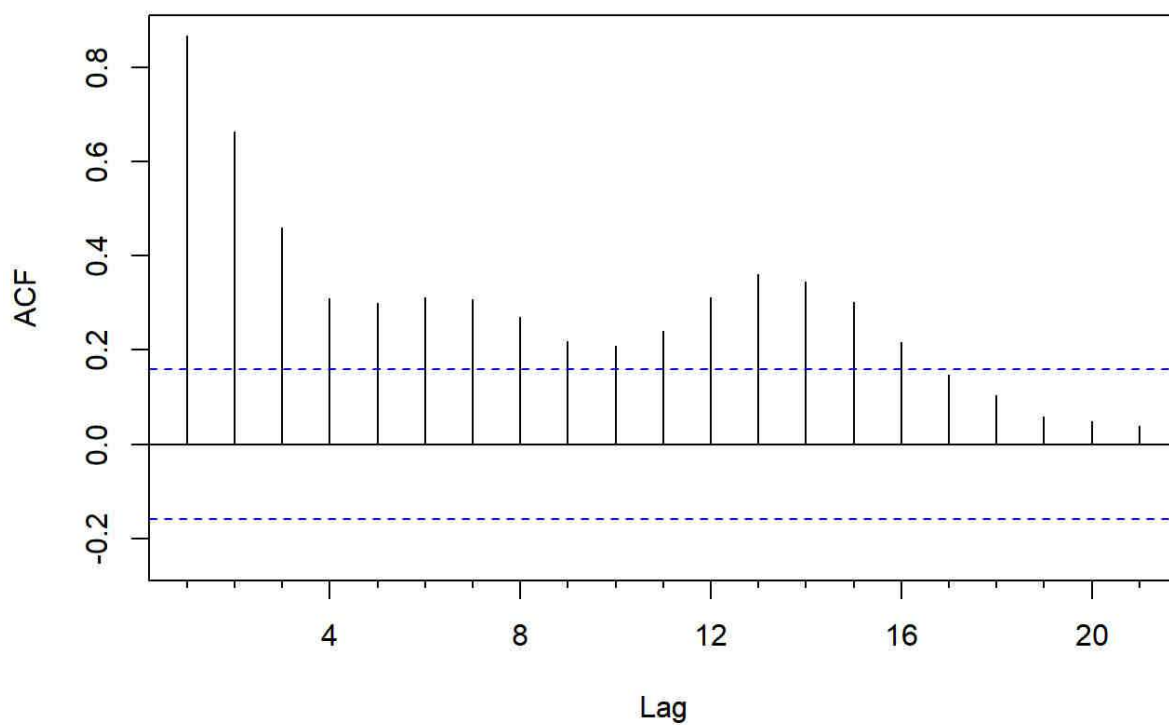
### Series residuos.lm



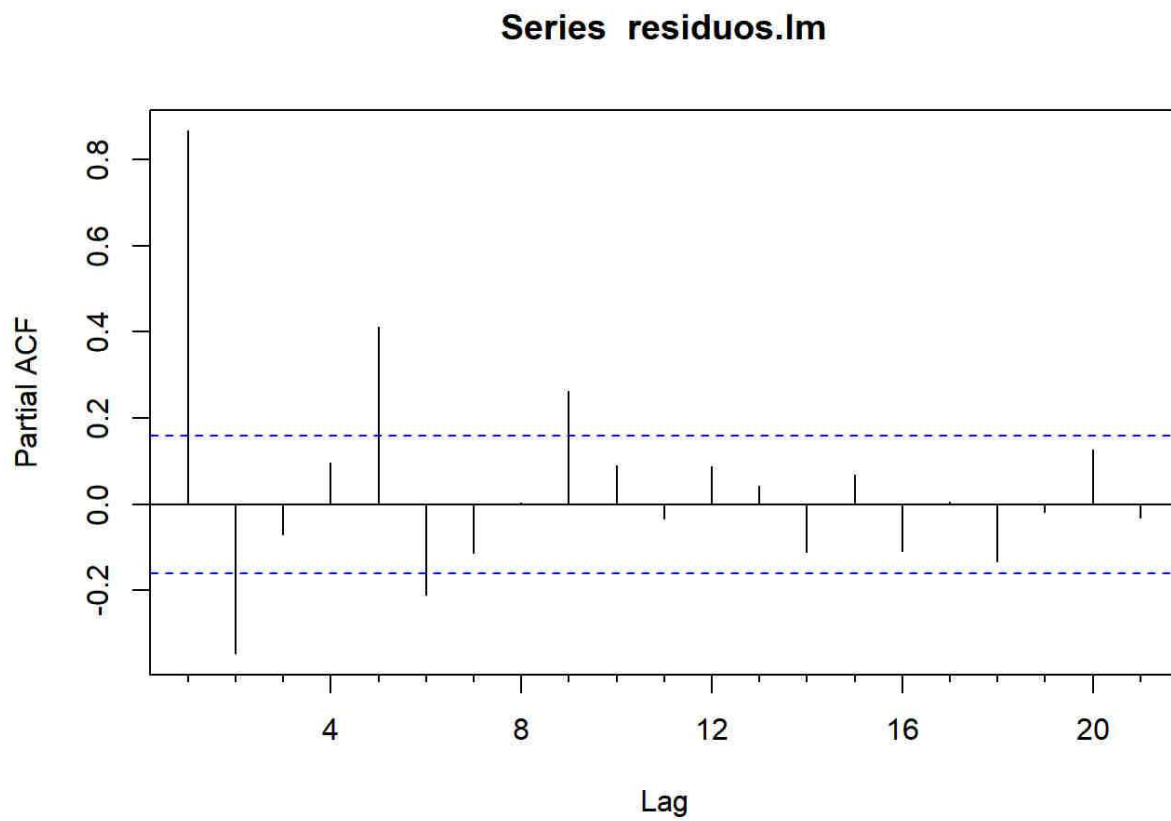
## 5.4.1 Correlograma con la Librería “forecast” (varias opciones)

```
# 1. help Acf (con "mayúsculas") para ver mejoras con respecto a acf
Acf(residuos.lm)
```

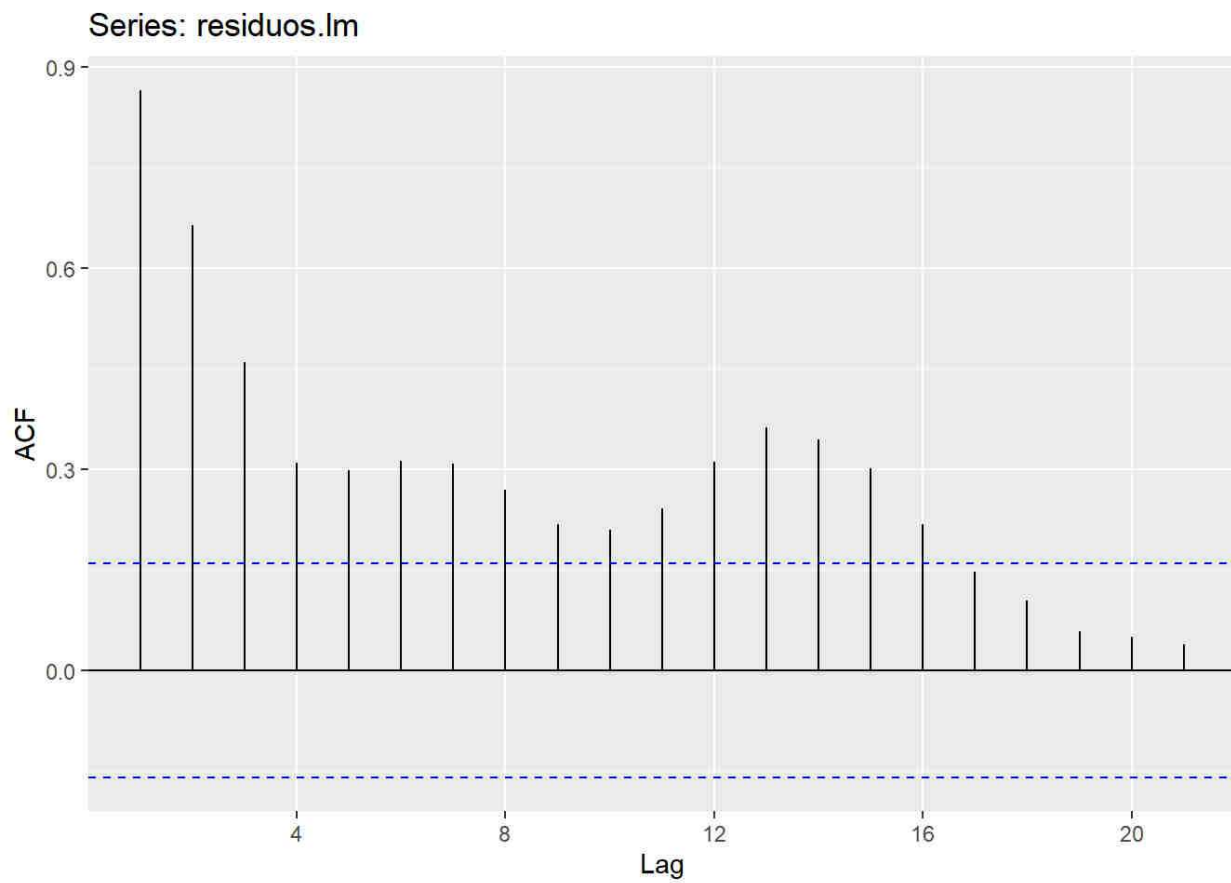
### Series residuos.lm



```
Pacf(residuos.lm)
```

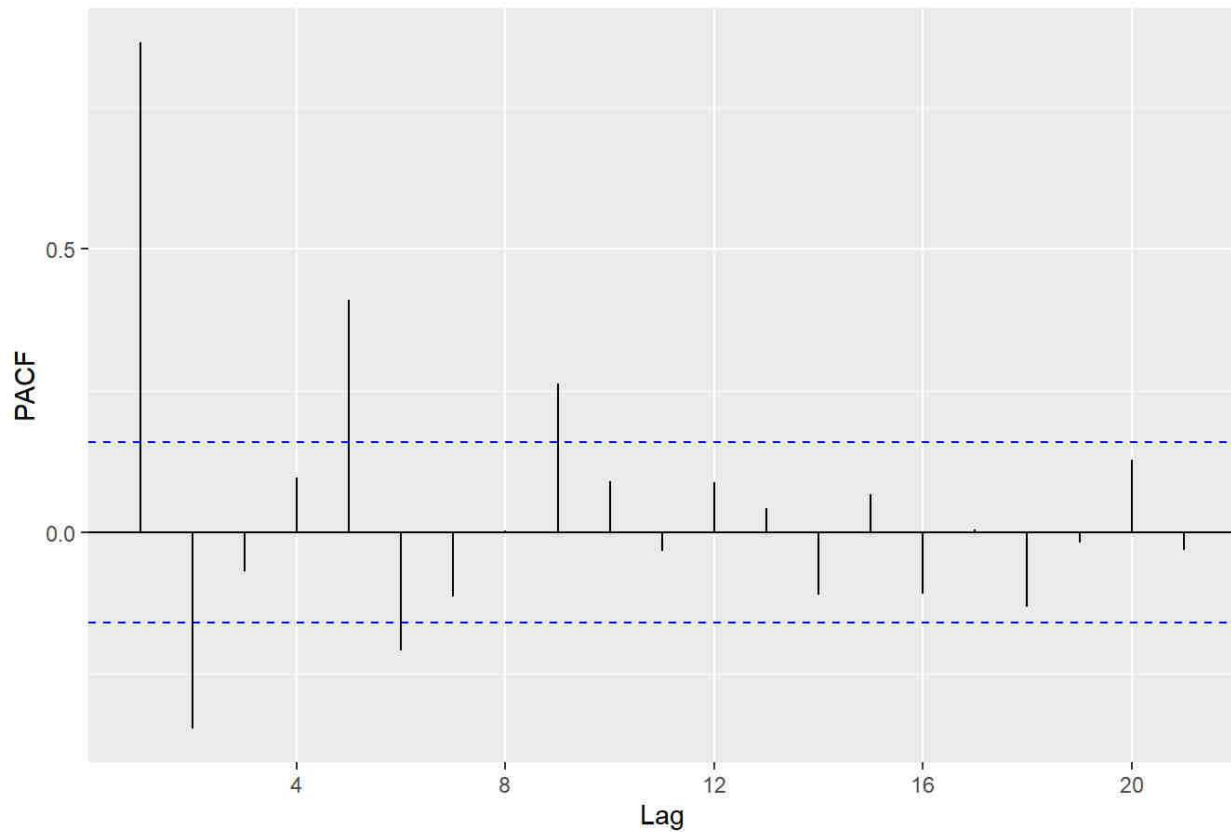


```
# 2. objeto ggplot  
ggAcf(residuos.lm)
```



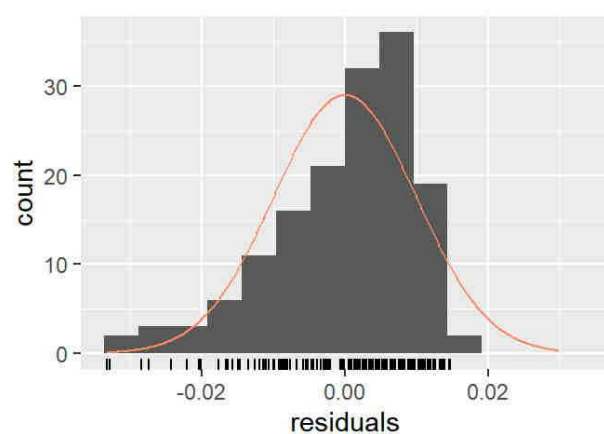
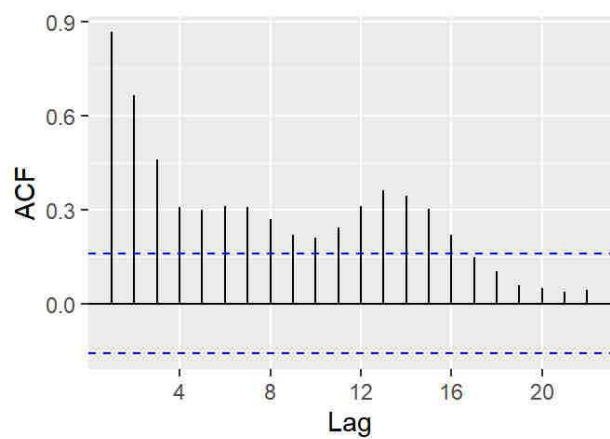
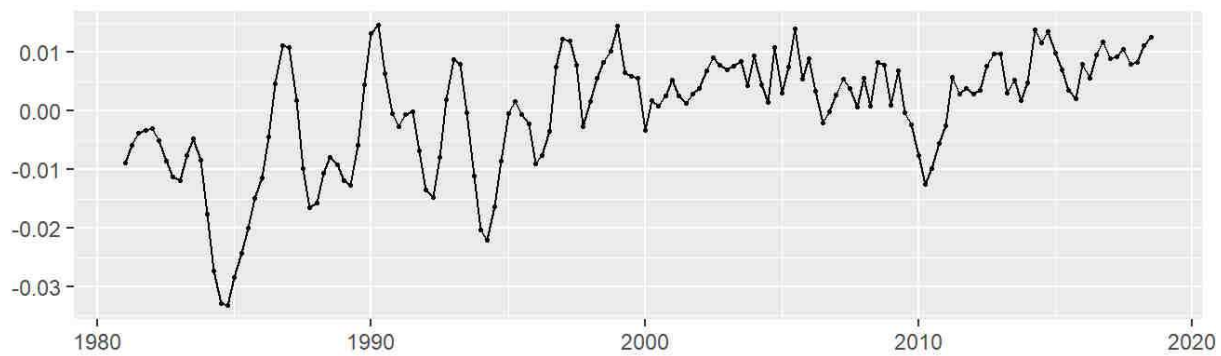
```
ggPacf(residuos.lm)
```

Series: residuos.lm

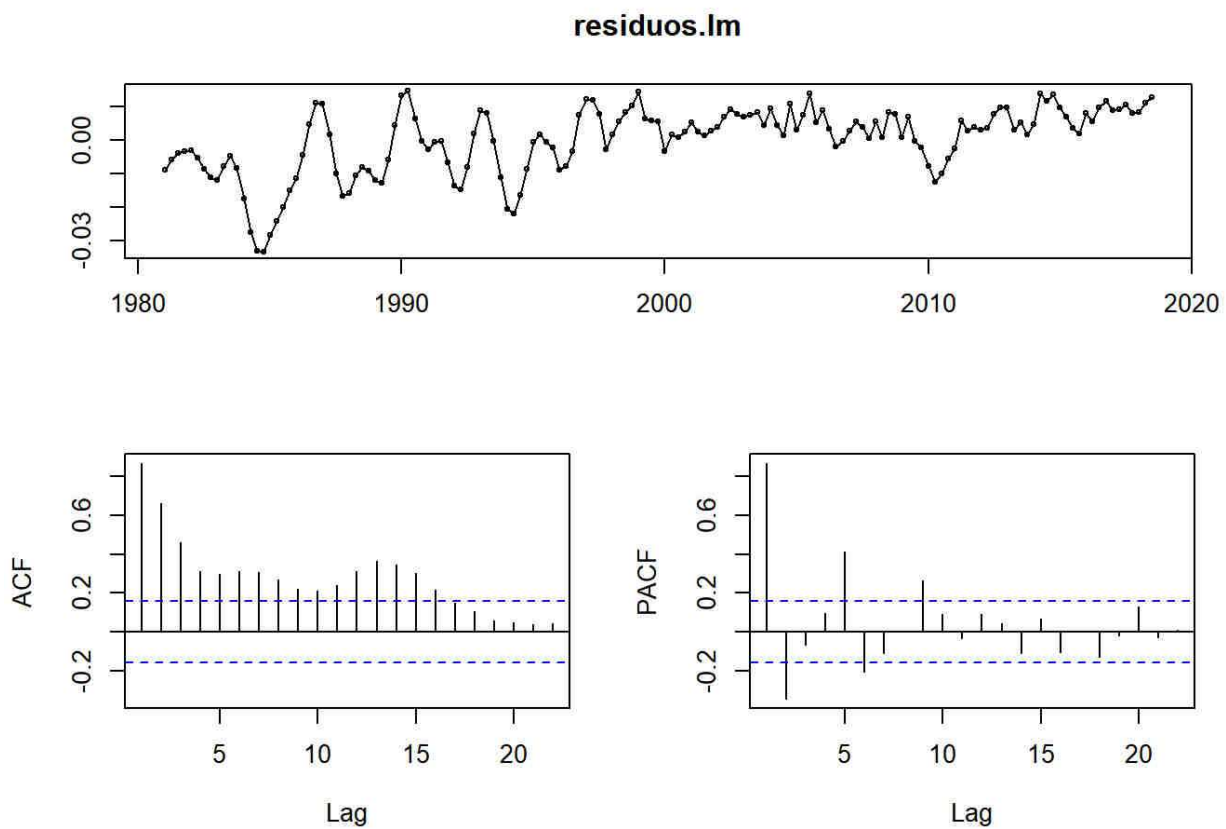


```
# 3. Gráfico residuos, acf, histograma  
checkresiduals(modelo.lm, test=FALSE)
```

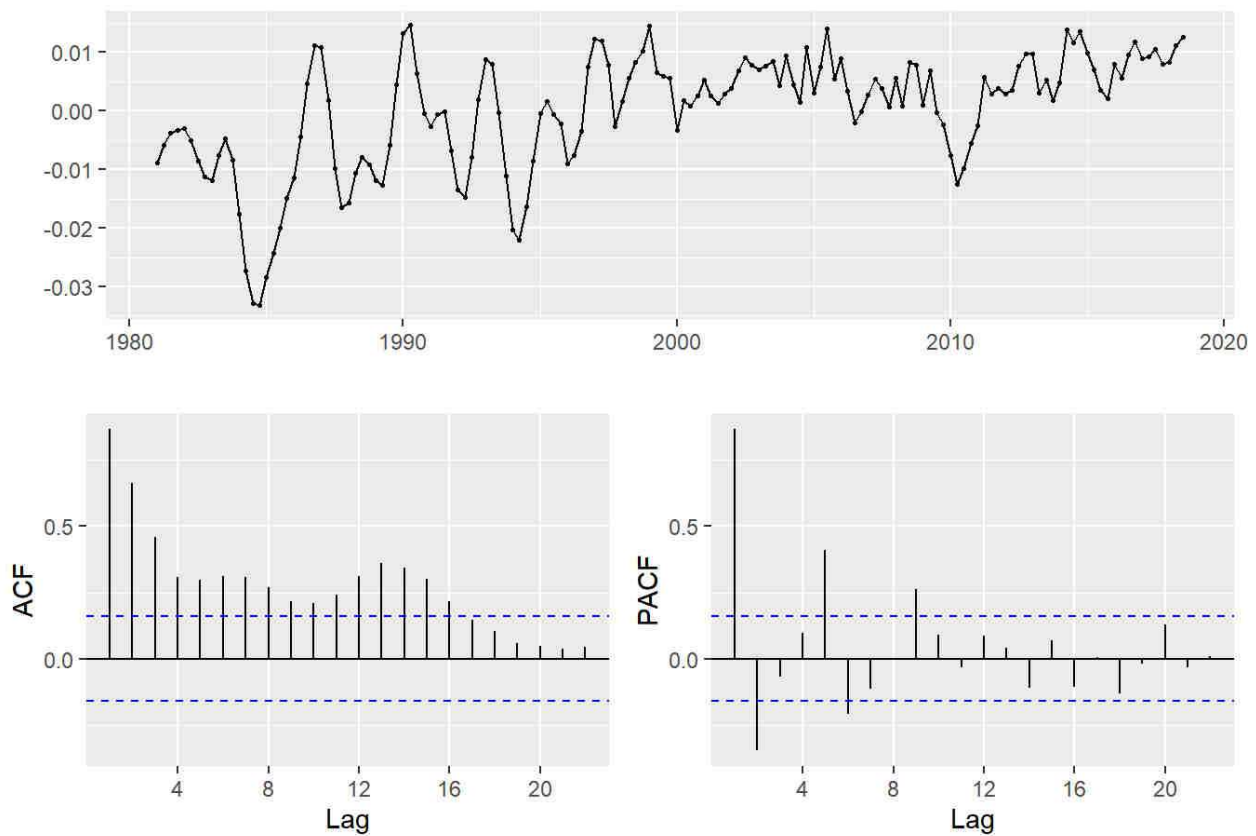
Residuals



```
# 4. Gráfico residuos, acf, pacf
tsdisplay(residuos.lm)
```



```
ggtsdisplay(residuos.lm) # Objeto ggplot
```



## 6 Paso 4: Estimación del modelo con autocorrelación en el error

```
# Paso 4: Estimación del modelo con autocorrelación----  
#=====
```

### 6.1 (1): Estimador MCO pero con un estimador consistente (Newey-West) de la matriz var-cov de los coeficientes MCO

Cuando cuando detectamos autocorrelación en los residuos todavía podemos seguir utilizando MCO para estimar los coeficientes del modelo porque los estimadores MCO siguen siendo insesgados (aunque no eficientes). El problema es que cuando existe autocorrelación la estimación  $X'X^{-1}$  no es ya la verdadera matriz de varianzas y covarianzas del estimador MCO, por lo que si decidimos utilizar MCO tendremos que utilizar estimador consistente (Newey-West) para la verdadera matriz var-cov de los coeficientes MCO

Esto se puede hacer con la librería `sandwich`

```
#  
  
modelo.lm.NW<-lmtest::coeftest(modelo.lm, vcov = sandwich::NeweyWest(modelo.lm))  
  
modelo.lm.NW <- lmtest::coeftest(modelo.lm, vcov = sandwich::NeweyWest(modelo.lm))  
print(modelo.lm.NW)
```

```
##  
## t test of coefficients:  
##  
##           Estimate Std. Error t value Pr(>|t|)  
## (Intercept) -0.0177127  0.0034186 -5.1812  7.05e-07 ***  
## T4pib.ts     1.2434860  0.0654062 19.0117 < 2.2e-16 ***  
## ---  
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

En esta tabla se pueden comprobar que los coeficientes estimados son los mismos, pero en la segunda las desviaciones típicas de los coeficientes se han estimado consistentemente

```
stargazer(modelo.lm, modelo.lm.NW, type="text")
```



```
##
## =====
##                      Dependent variable:
##          -----
##                      T4ocu.ts
##                      OLS                      coefficient
##                      test
##                      (1)                      (2)
##          -----
## T4pib.ts                      1.243***          1.243***
##                      (0.036)          (0.065)
##
## Constant                      -0.018***          -0.018***
##                      (0.001)          (0.003)
##
##          -----
## Observations                      151
## R2                      0.888
## Adjusted R2                      0.887
## Residual Std. Error          0.010 (df = 149)
## F Statistic          1,182.905*** (df = 1; 149)
## =====
## Note:                      *p<0.1; **p<0.05; ***p<0.01
```

```
# _____ #
```

## 6.2 (2) Mínimos Cuadrados Generalizados Factible - errores AR(1)

Una segunda Opción consiste en dejar de usar MCO y utilizar otro estimador que además de insesgado sea eficiente (mínima varianza). Este estimador es el de **Mínimos Cuadrados Generalizados** que lo que hace es estimar de alguna manera la estructura de autocorrelación que aparece en la matriz de Var-Cov del término de error

Cuando los errores se suponen que están autocorrelacionados de orden uno AR(1) pueden utilizarse dos paquetes para estimar por Mínimo Cuadrados Generalizados, con las librerías `orcutt` y `prais`

```
# Cochrane-Orcutt (perdemos la primera observación)
orcutt::cochrane.orcutt(modelo.lm)
```

```
## Cochrane-orcutt estimation for first order autocorrelation
##
## Call:
## lm(formula = modelo, data = datokun, na.action = NULL)
##
## number of interaction: 10
## rho 0.937181
##
## Durbin-Watson statistic
## (original): 0.25147 , p-value: 6.2e-28
## (transformed): 1.25404 , p-value: 1.619e-06
##
## coefficients:
## (Intercept) T4pib.ts
## -0.005964 0.874841
```

```
# Prais-Winsten
prais::prais_winsten(modelo.lm, data=datokun)
```

```
## Iteration 0: rho = 0
## Iteration 1: rho = 0.8757
## Iteration 2: rho = 0.9268
## Iteration 3: rho = 0.9348
## Iteration 4: rho = 0.9358
## Iteration 5: rho = 0.9359
## Iteration 6: rho = 0.9359
## Iteration 7: rho = 0.9359
## Iteration 8: rho = 0.9359
```

```
##
## Call:
## prais::prais_winsten(formula = modelo.lm, data = datokun)
##
## Coefficients:
## (Intercept)      T4pib.ts
##   -0.009827      0.887548
##
## AR(1) coefficient rho: 0.9359
```

```
#prais.winsten(modelo, data=datokun)
```

```
#_____#
```

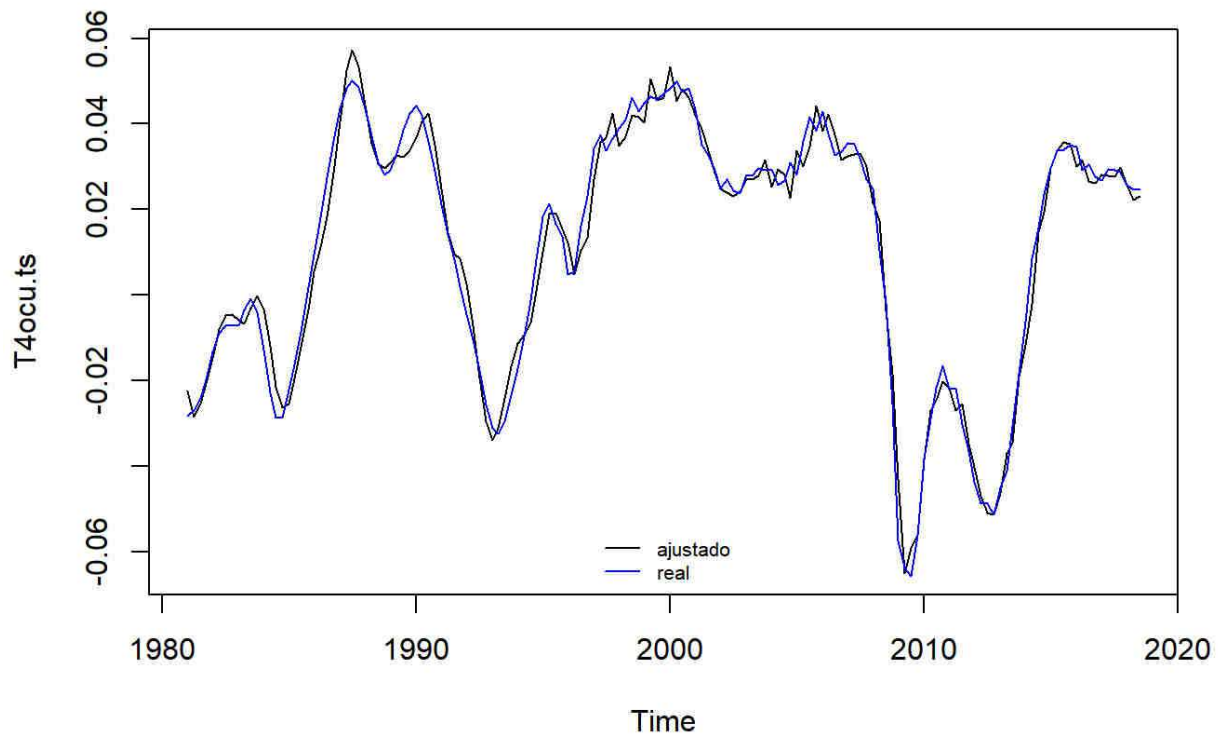
## 6.3 (3) Estimo modelo lineal con un proceso AR(1) en el término de error

Cuando la estructura autoregresiva del término de error es más compleja podemos estimarla suponiendo que esos errores siguen una estructura ARMA, comenzamos suponiendo estructura AR(1)

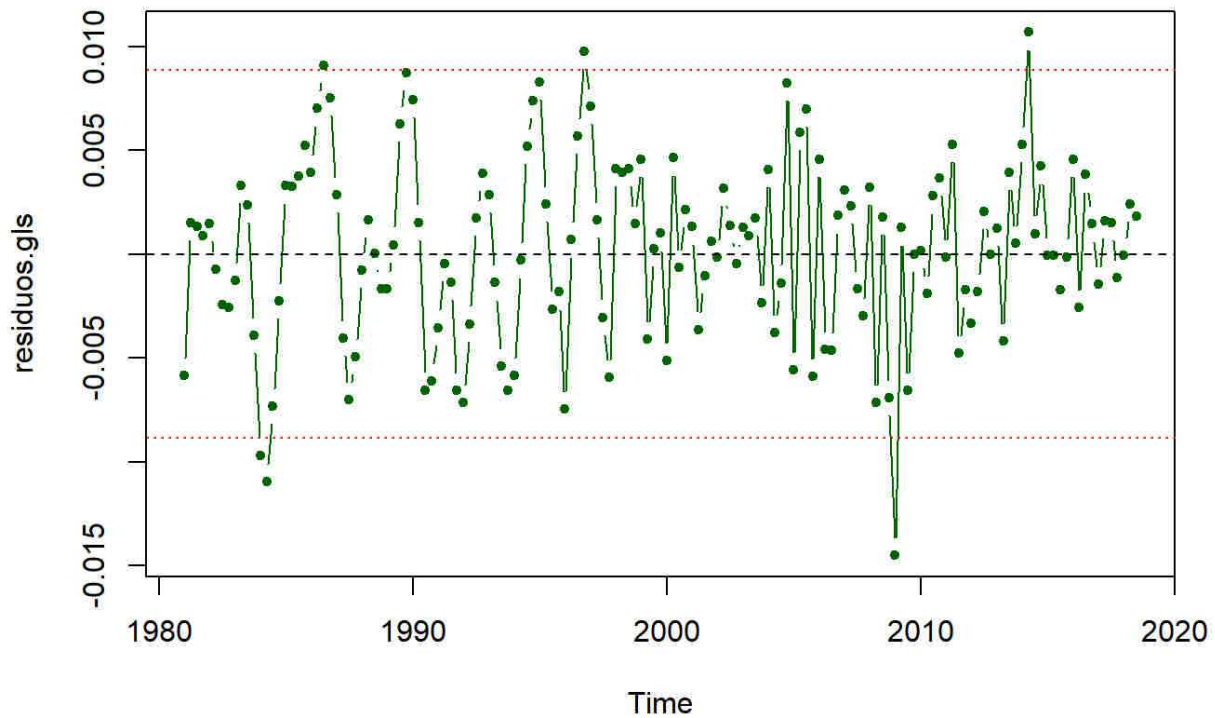
```
#modelo.gls <- arima(x = T4ocu.ts, xreg= T4pib.ts, order=c(1,0,0)) # Base de R
modelo.gls <- Arima(T4ocu.ts, xreg= T4pib.ts, order=c(1,0,0))      # Forecast
print(modelo.gls)
```

```
## Series: T4ocu.ts
## Regression with ARIMA(1,0,0) errors
##
## Coefficients:
##           ar1  intercept    xreg
##       0.9415   -0.0097  0.8834
## s.e.  0.0283    0.0057  0.0662
##
## sigma^2 estimated as 1.99e-05: log likelihood=603.45
## AIC=-1198.91  AICc=-1198.63  BIC=-1186.84
```

```
# valor ajustado: T4ocu.ts-modelo.gls$residuals
plot(T4ocu.ts-modelo.gls$residuals, ylab="T4ocu.ts", type="l")
lines(T4ocu.ts,col="blue")
legend("bottom",legend = c("ajustado", "real"), lty=c(1), cex=0.65, lwd=c(1,1),
      bty="n",
      col=c("black","blue"))
```

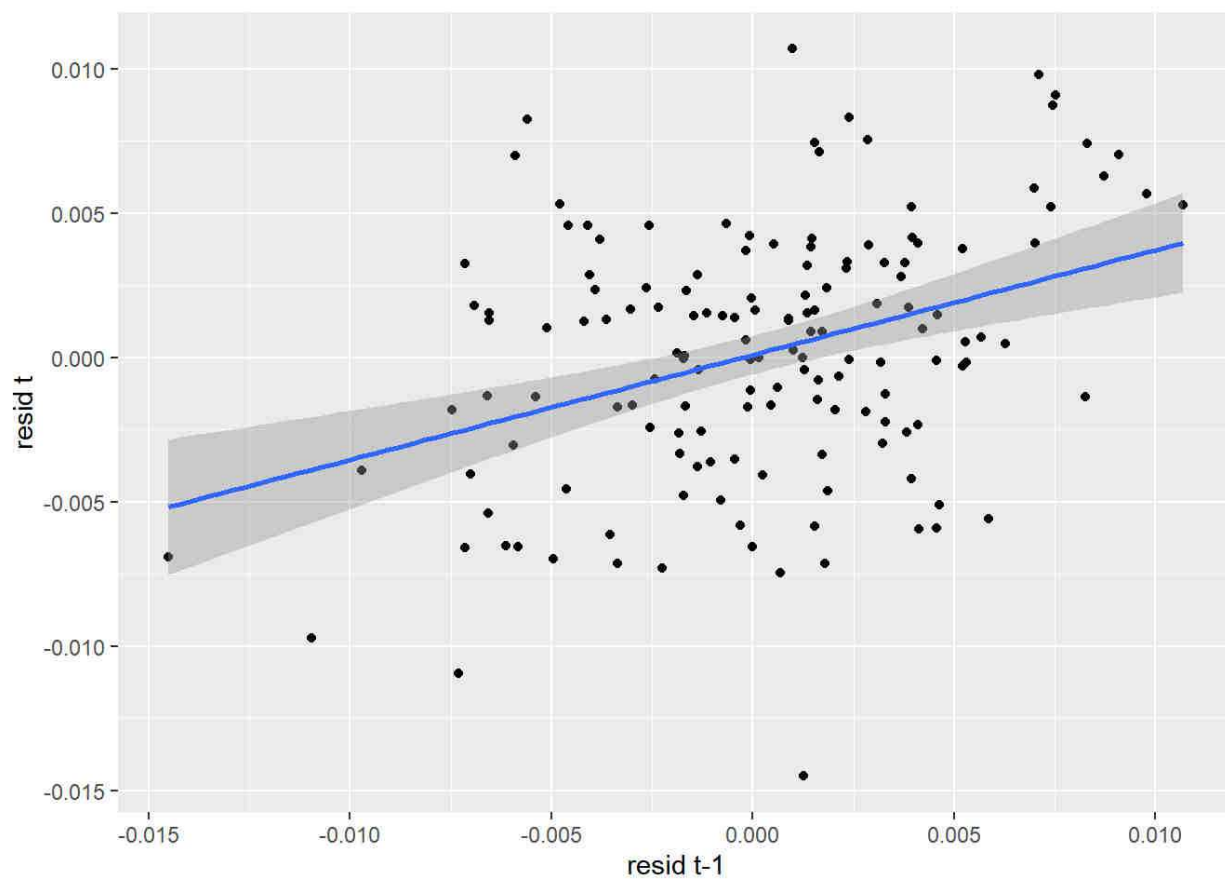


```
#### Diagnósis de los residuos
residuos.gls <- residuals(modelo.gls)
# _____ #
plot(residuos.gls, type="b", col="darkgreen",pch=20)
abline(h=0,lty=2, col="grey2" )
abline(h=2*sd(residuos.gls),lty=3, col="red" )
abline(h=-2*sd(residuos.gls),lty=3, col="red" )
```



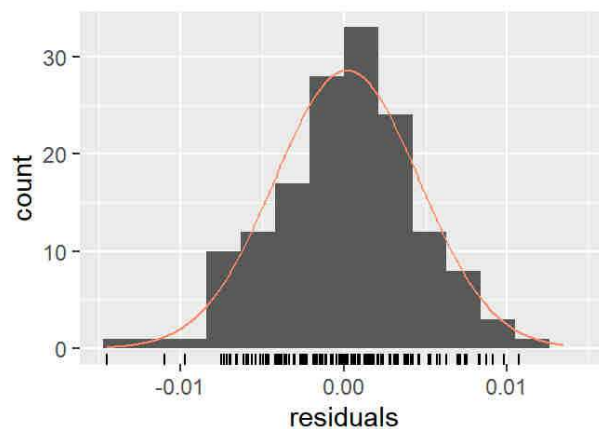
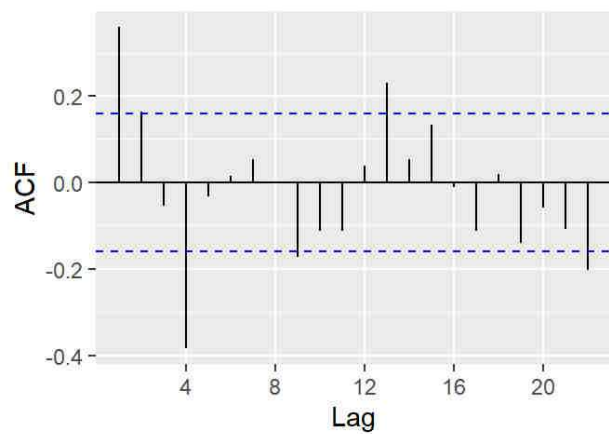
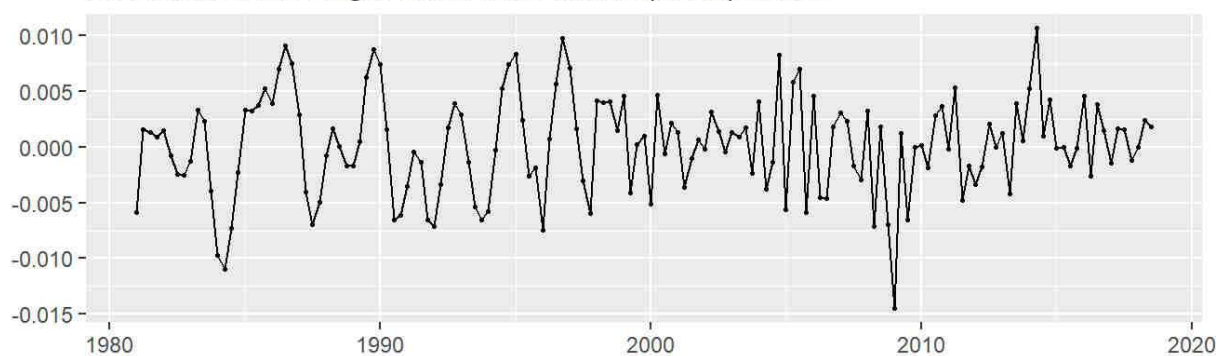
```
residuos_plot <- ts.intersect("resid_t=residuos.gls, "resid_t1"=lag(residuos.gls, k = 1), dframe=TRUE)
ggplot(data=residuos_plot)+
  geom_point(aes(x=resid_t1, y=resid_t))+
  xlab("resid t-1")+
  ylab("resid t")+
  geom_smooth(aes(x=resid_t1,y=resid_t), method = "lm")
```

```
## Don't know how to automatically pick scale for object of type ts. Defaulting to continuous.
## Don't know how to automatically pick scale for object of type ts. Defaulting to continuous.
```



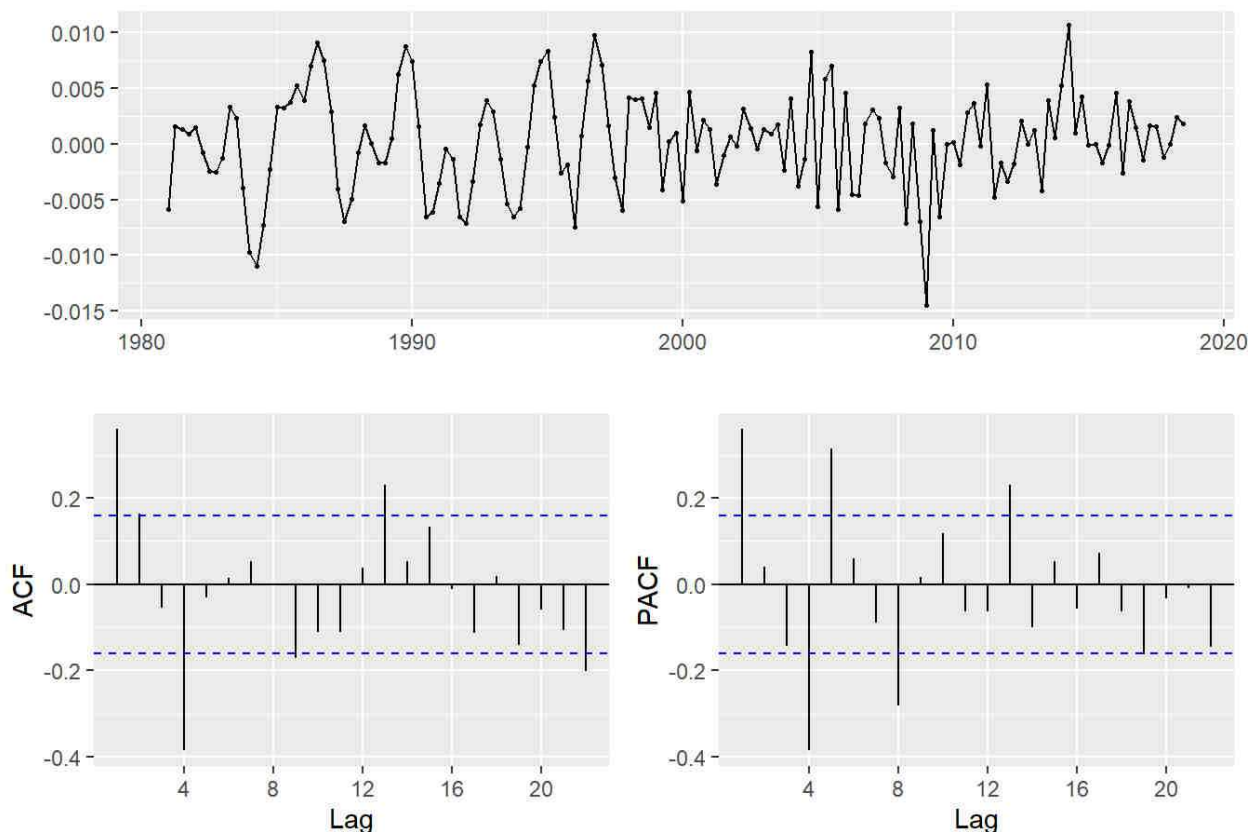
```
# Correlograma & Ljung-Box test
# Librería forecast (permite calcular el contraste Ljung-Box)
checkresiduals(modelo.gls) # Ljung-Box test
```

Residuals from Regression with ARIMA(1,0,0) errors



```
##
## Ljung-Box test
##
## data: Residuals from Regression with ARIMA(1,0,0) errors
## Q* = 48.169, df = 5, p-value = 3.28e-09
##
## Model df: 3. Total lags used: 8
```

```
ggtsdisplay(residuos.gls) # Objeto ggplot
```



```
# _____ #
```

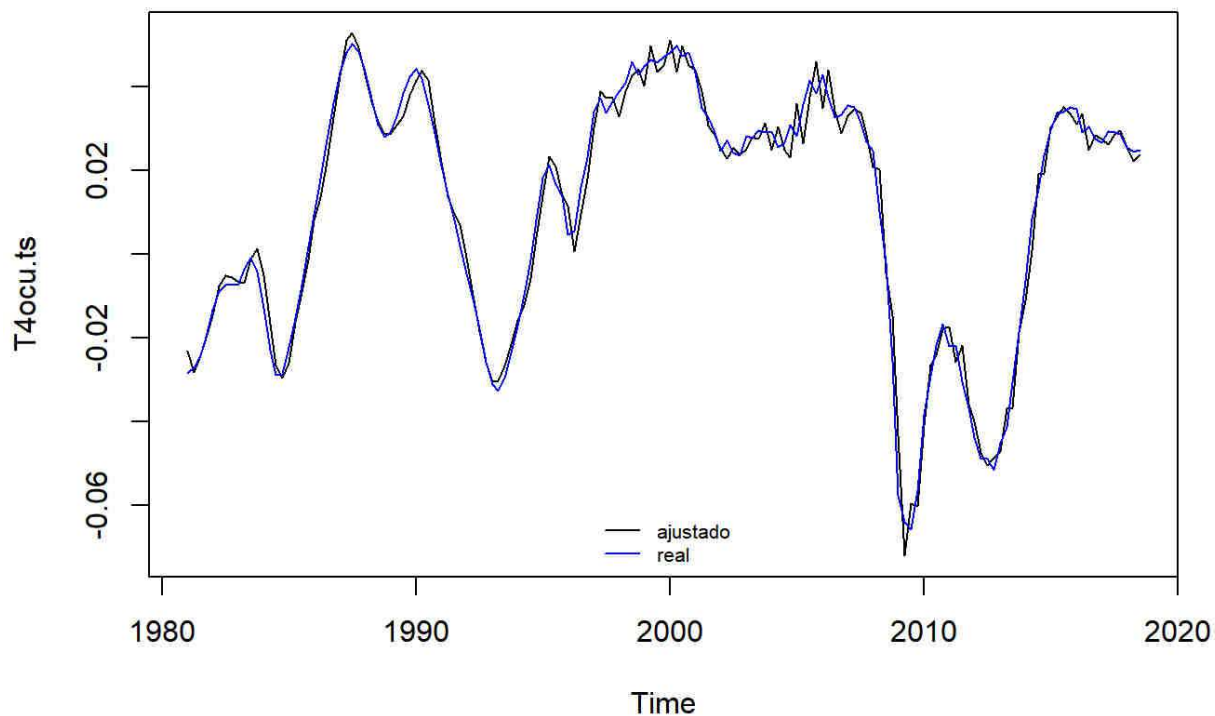
## 6.4 (4) Estimo modelo lineal con un proceso AR(2) en el término de error

Podemos especificar cualquier tipo de estructura autoregresiva, por ejemplo un AR(2)

```
#modelo.gls2 <- arima(x = T4ocu.ts, xreg= T4pib.ts, order=c(2,0,0)) # Base de R
modelo.gls2 <- Arima(T4ocu.ts, xreg= T4pib.ts, order=c(2,0,0)) # Forecast
print(modelo.gls2)
```

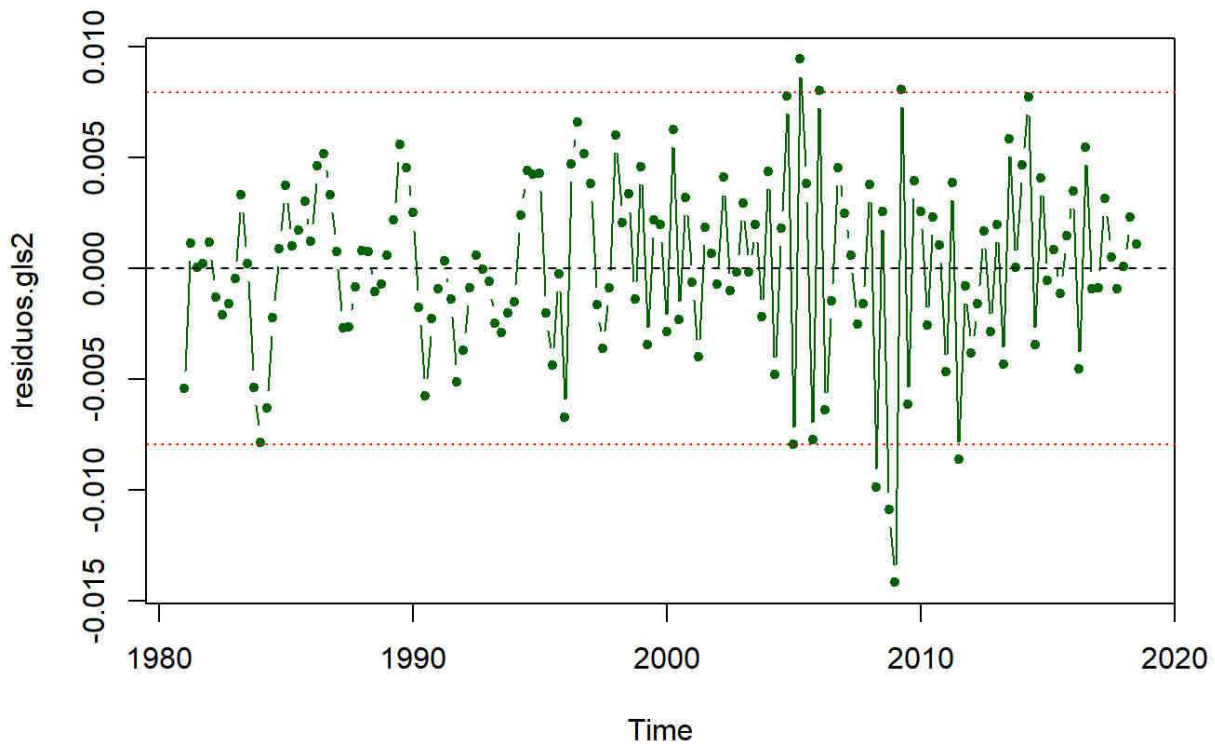
```
## Series: T4ocu.ts
## Regression with ARIMA(2,0,0) errors
##
## Coefficients:
##          ar1      ar2  intercept    xreg
##       1.4651  -0.5199   -0.0036  0.6146
## s.e.  0.0939   0.0902    0.0061  0.1084
##
## sigma^2 estimated as 1.616e-05:  log likelihood=619.14
## AIC=-1228.28   AICc=-1227.87   BIC=-1213.19
```

```
# Valor ajustado: T4ocu.ts-modelo.gls2$residuals
plot(T4ocu.ts-modelo.gls2$residuals, ylab="T4ocu.ts", type="l")
lines(T4ocu.ts,col="blue")
legend("bottom",legend = c("ajustado", "real"),lty=c(1), cex=0.65, lwd=c(1,1),
      bty="n",
      col=c("black","blue"))
```



```
# Diagn sis de los residuos
residuos.gls2 <- residuals(modelo.gls2)

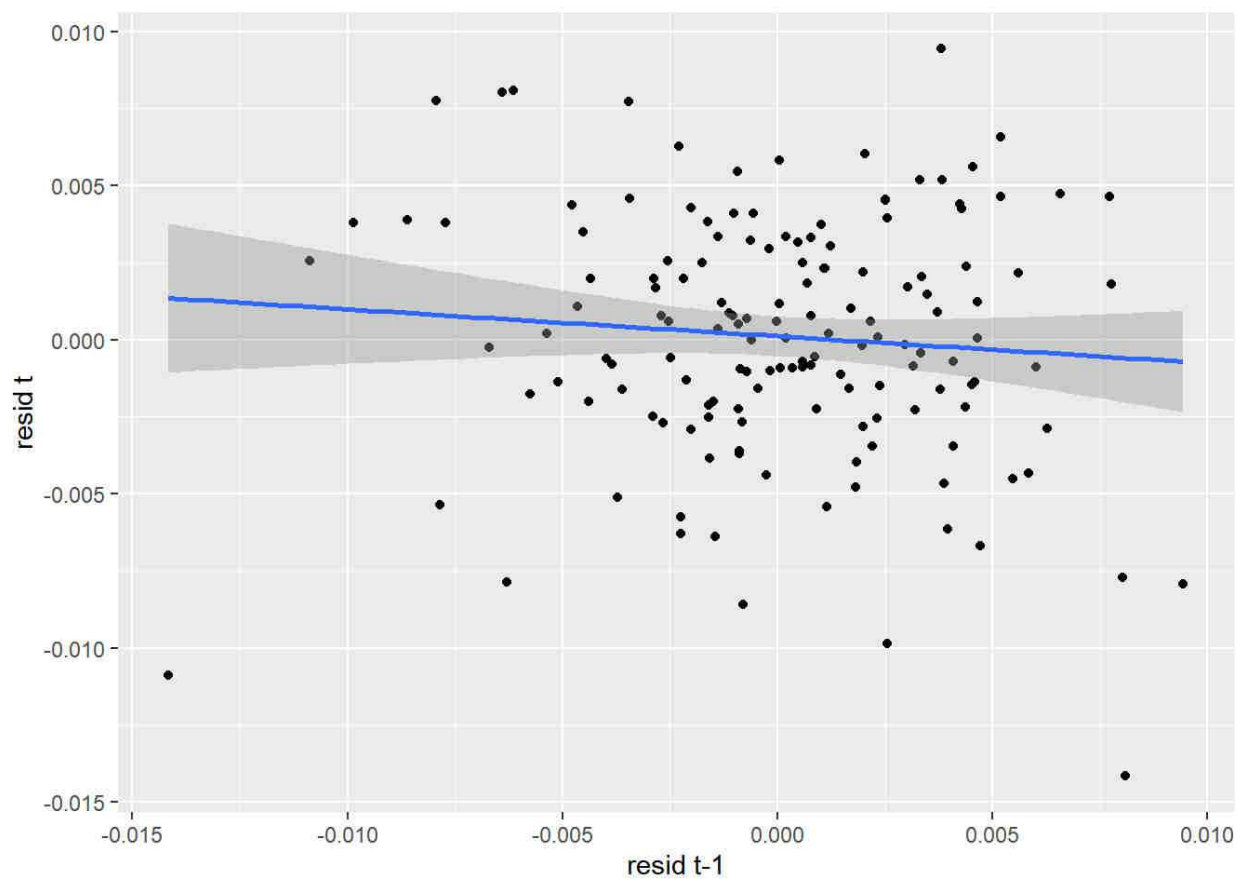
plot(residuos.gls2, type="b", col="darkgreen",pch=20)
abline(h=0,lty=2, col="grey2" )
abline(h=2*sd(residuos.gls2),lty=3, col="red" )
abline(h=-2*sd(residuos.gls2),lty=3, col="red" )
```



```
# ggplot (dibuja datos data.frame)
residuos_plot2 <- ts.intersect("resid_t"=residuos.gls2, "resid_t1"=lag(residuos.gls2, k = 1), dframe
=TRUE)
ggplot(data=residuos_plot2)+
  geom_point(aes(x=resid_t1,y=resid_t))+
  xlab("resid t-1")+
  ylab("resid t")+
  geom_smooth(aes(x=resid_t1,y=resid_t),method = "lm")
```

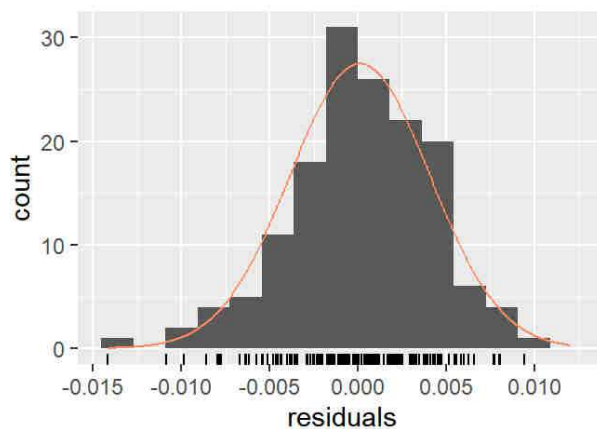
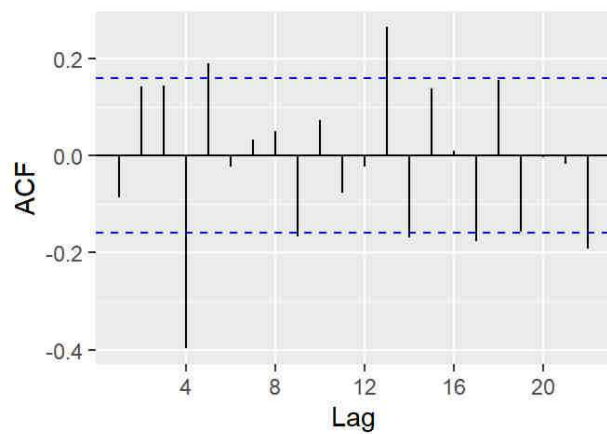
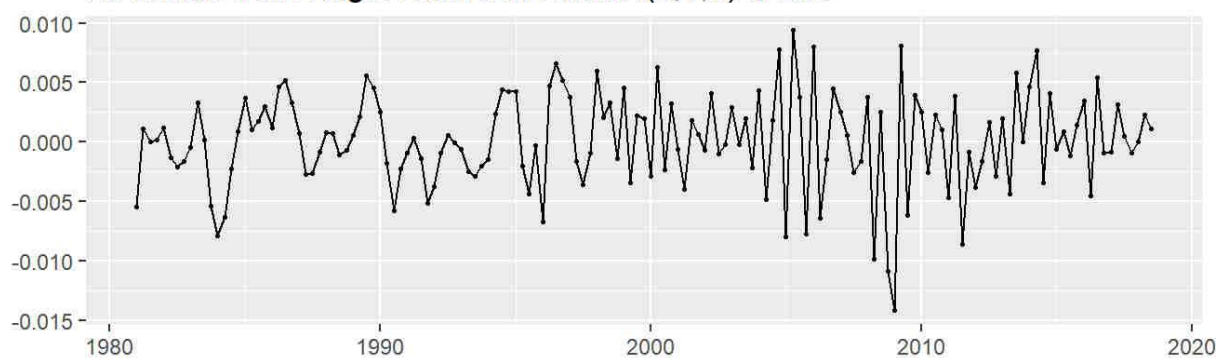
```
## Don't know how to automatically pick scale for object of type ts. Defaulting to continuous.
## Don't know how to automatically pick scale for object of type ts. Defaulting to continuous.
```





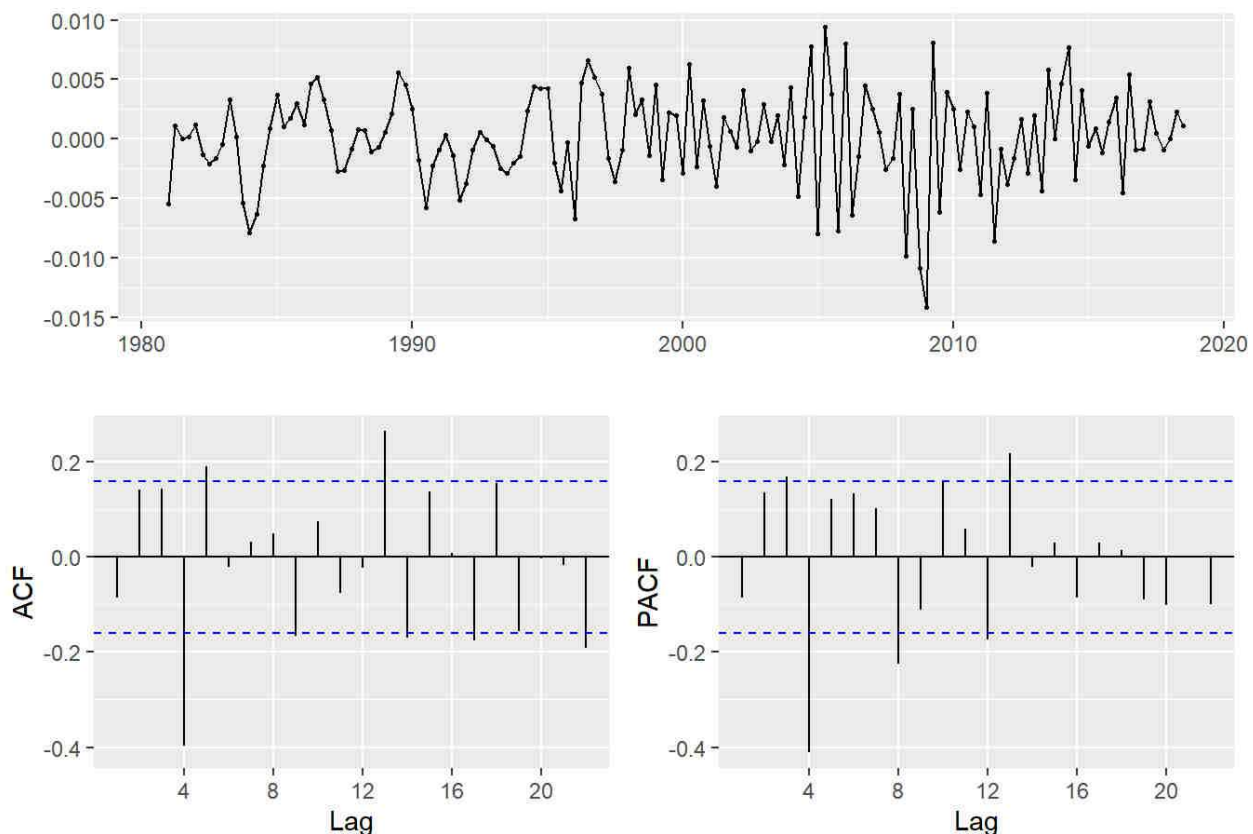
```
# Correlograma & Ljung-Box test
# Librería forecast (permite calcular el contraste Ljung-Box)
checkresiduals(modelo.gls2, test="LB") # Ljung-Box test
```

Residuals from Regression with ARIMA(2,0,0) errors



```
##
## Ljung-Box test
##
## data: Residuals from Regression with ARIMA(2,0,0) errors
## Q* = 38.526, df = 4, p-value = 8.727e-08
##
## Model df: 4. Total lags used: 8
```

```
ggtsdisplay(residuos.gls2) # Objeto ggplot
```



```
stargazer(modelo.gls, modelo.gls2 , type="text")
```

```
##
## % Error: Unrecognized object type.
## % Error: Unrecognized object type.
```

```
# _____ #
```

## 6.5 (5) Estimo modelo lineal con un proceso `auto.arima()` en el término de error

El caso más general es cuando dejo que sea el propio ajuste el que determine la estructura autoregresiva del término de error

```

modelo.auto <- auto.arima(T4ocu.ts, xreg= T4pib.ts)
modelo.auto <- auto.arima(T4ocu.ts, xreg= T4pib.ts, seasonal=FALSE)
print(modelo.auto)

```

```

## Series: T4ocu.ts
## Regression with ARIMA(4,1,2) errors
##
## Coefficients:
##          ar1          ar2          ar3          ar4          ma1          ma2          drift          xreg
##          0.2445         -0.3618         0.1686         -0.3249         0.1247         0.8603         2e-04         0.7763
## s.e.      0.1183         0.0974         0.1102         0.0977         0.0813         0.0452         4e-04         0.1458
##
## sigma^2 estimated as 1.191e-05:  log likelihood=640.14
## AIC=-1262.29   AICc=-1261    BIC=-1235.19

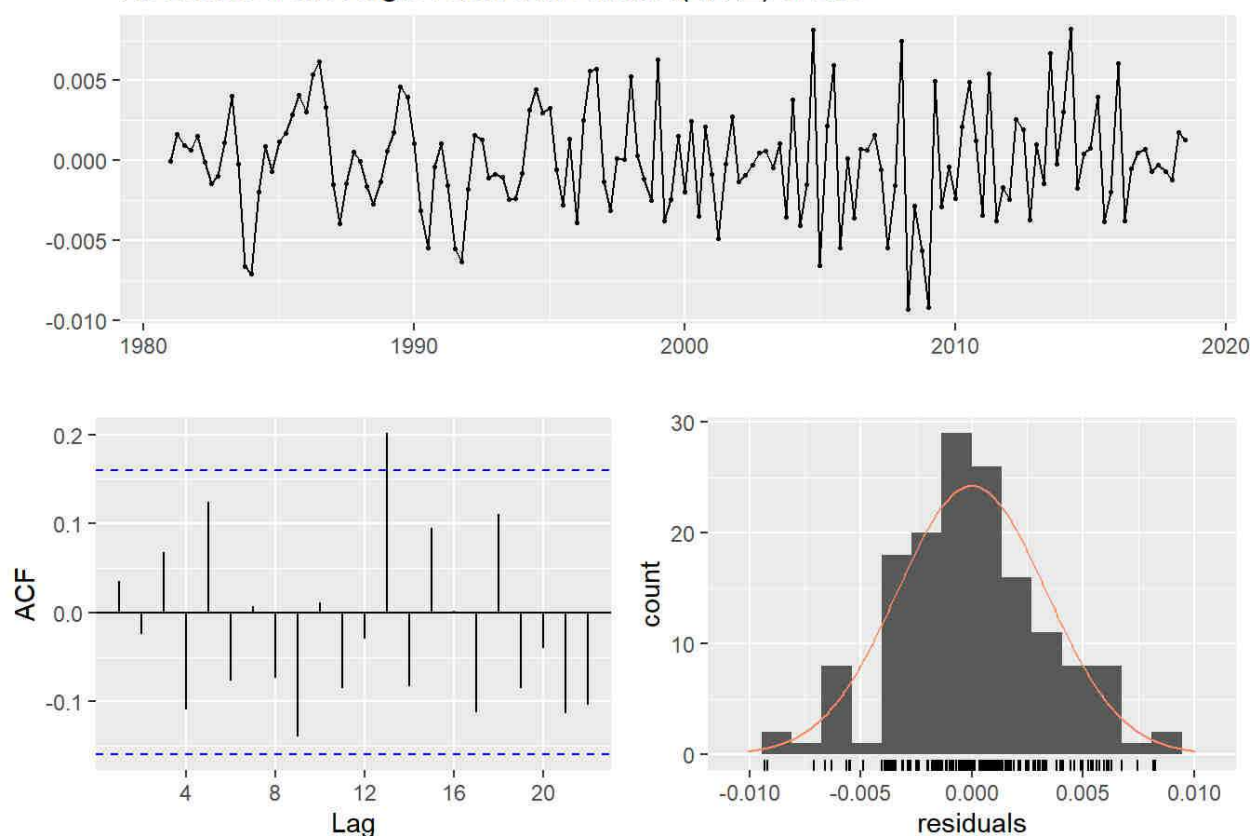
```

```

# Correlograma & Ljung-Box test
# Librería forecast (permite calcular el contraste Ljung-Box)
residuos.auto <- residuals(modelo.auto)
checkresiduals(modelo.auto, test="LB")

```

### Residuals from Regression with ARIMA(4,1,2) errors



```

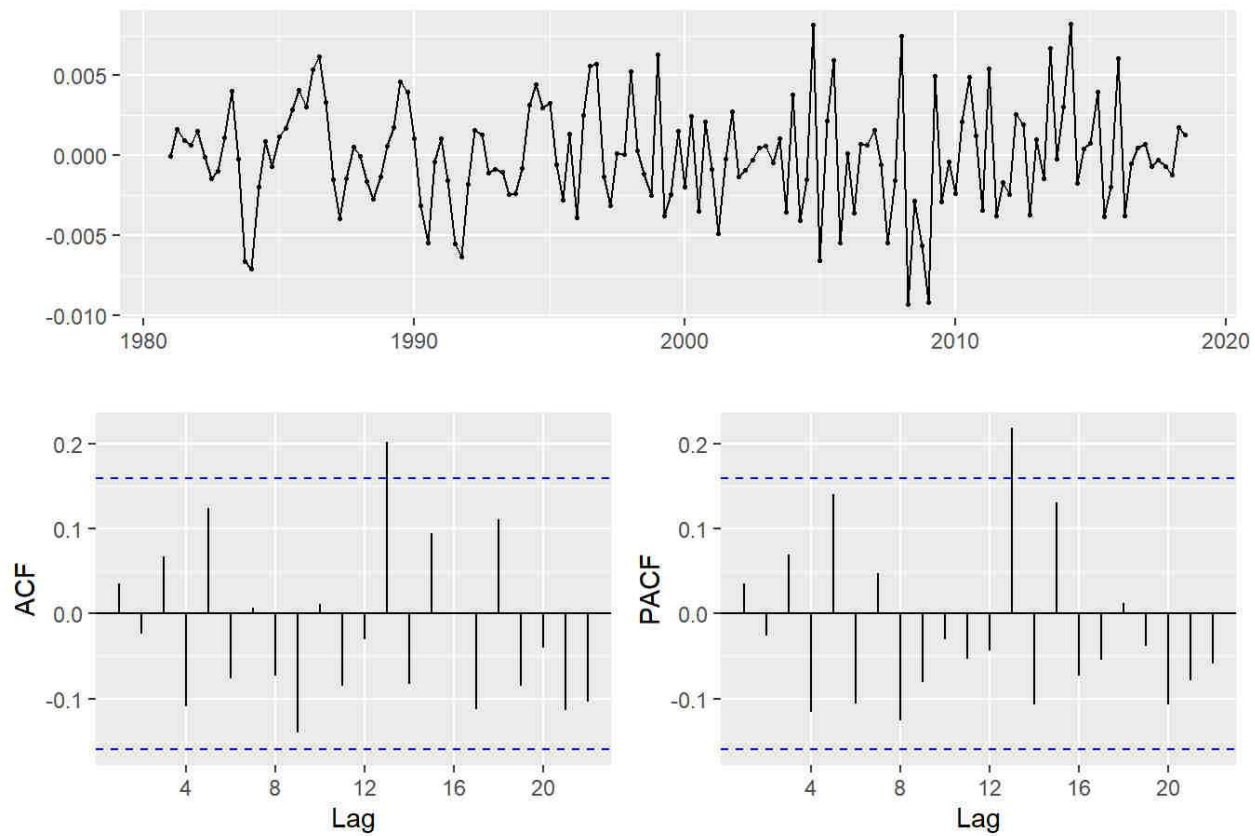
##
## Ljung-Box test
##
## data: Residuals from Regression with ARIMA(4,1,2) errors
## Q* = 11.506, df = 3, p-value = 0.009284
##
## Model df: 8. Total lags used: 11

```

```

ggtsdisplay(residuos.auto)

```



#

#



Proyecto de innovación  
“Innova Docencia” de la UCM  
Convocatoria 2018-2019  
Proyecto número 275

**R-adaptación de la asignatura de métodos econométricos en economía y finanzas del  
Grado de Estadística Aplicada**

Facultad de Estudios Estadísticos  
Departamento de Economía Aplicada, Pública y Política

**Título**

***Práctica 12: El Índice de Confianza del Consumidor como indicador  
adelantado del PIB***

**Correspondiente al**

**Capítulo 5 MODELOS CON DATOS DE SERIES TEMPORALES II: Autocorrelación,  
indicadores adelantados y causalidad en sentido de Granger**

- Modelos de regresión con **series temporales estacionarias**: autocorrelación
- Contrastes de autocorrelación
- El estimador de Mínimos Cuadrados Generalizados (MCG)  
***Práctica 11: Autocorrelación en La Ley de Okun***
- Indicadores adelantados: Causalidad en sentido de Granger y  
correlación contemporánea  
***Práctica 12: El Índice de Confianza del Consumidor como indicador  
adelantado del PIB***

# Modelos de regresión con series temporales estacionarias: Indicadores Adelantados y Causalidad en sentido de Granger

## Asignatura Métodos Econométricos en Economía y Finanzas

Profesores  
Loren Escot, Alicia Pérez-Alonso y Julio E. Sandubete

Dpto Economía Aplicada, Pública y Política  
Facultad Estudios Estadísticos - Universidad Complutense de Madrid

Curso académico 2018/2019

### Abstract

En esta práctica se muestra cómo detectar cuando dos series temporales presentan correlación no contemporánea, de forma que una de ellas pueda utilizarse como indicador adelantado de la otra. En este sentido se presenta el **test de Granger de Causalidad** entre series temporales.

Como ejemplo nos preguntaremos si el **Indicador de Confianza del Consumidor** puede considerarse un buen indicador adelantado del **PIB**

- 1 Librerías necesarias
- 2 Carga de datos
  - 2.1 Leo los datos del PIB
  - 2.2 Carga de datos de indicadores Adelantados: Indicadores de Confianza del Consumidor
    - 2.2.1 Descarga del **ICC** desde el Eurostat
    - 2.2.2 Dibujo las dos series ¿cómo interpretar el ICC?
- 3 Transformo todas las series para que tengan la misma periodicidad (Trimestral)
- 4 Modelo Inicial: Asumiendo Correlaciones contemporáneas
  - 4.1 Representación gráfica de las Dos series
  - 4.2 Análisis de la Estacionariedad en media y Cointegración
  - 4.3 Estimo el modelo lineal MCO
- 5 Correlación NO contemporánea e indicadores adelantados
  - 5.1 Test de causalidad de Granger
- 6 Estimación del Modelo con indicador adelantado
  - 6.1 Predicción con los los indicadores adelantados

```
#=====#
#   File: t3granger.R
#   This script was written for R.3.4.3
#   First version: 16abr2018
#   Last version: 16abr2018
#   Last revisor:
#_____#
#   Description: INDICADORES ADELANTADOS
#               1. Test de causalidad de Granger
#
#   Packages: lmtest, tempdisagg
#   Input:
#   1) okundat.Rdata  [DATOS EMPLEO EQUIVALENTE];
#   2) Banco de Espa?a: ICC-?ndice de Confianza del Consumidor
#   Cuadro 1.1: https://www.bde.es/webbde/es/estadis/infoest/si_1_1.pdf
#   Inicio > Estad?sticas > S?ntesis de indicadores > 1. Espa?a >
#   > 1.1 - Indicadores econ?micos generales Ocultar informaci?n > Series temporales del cuadro 1.1
#=====#
```

# 1 Librerías necesarias

```
# Librerías necesarias ----
```

```
dev.off()                                # Cierra gr?ficos que podamos tener abiertos
```

```
## null device
##          1
```

```
rm(list=ls())                            # Elimina todos los objetos que tengamos en la memoria
#install.packages("lmtest")              # S?lo necesitas instalarlo la primera vez
#install.packages("tempdisagg")
#install.packages("forecast")

suppressMessages(library(lmtest))        # Test de Causalidad de Granger: grangertest(x,
y,..)
suppressMessages(library(tempdisagg))    # ta(): agrega series; td(): trimestraliza/mensual
  iza series anuales
suppressMessages(library(forecast))      # gCcf():
suppressMessages(library(stargazer))     # PAra sacar tablas de resultados

#_____#
```

## 2 Carga de datos

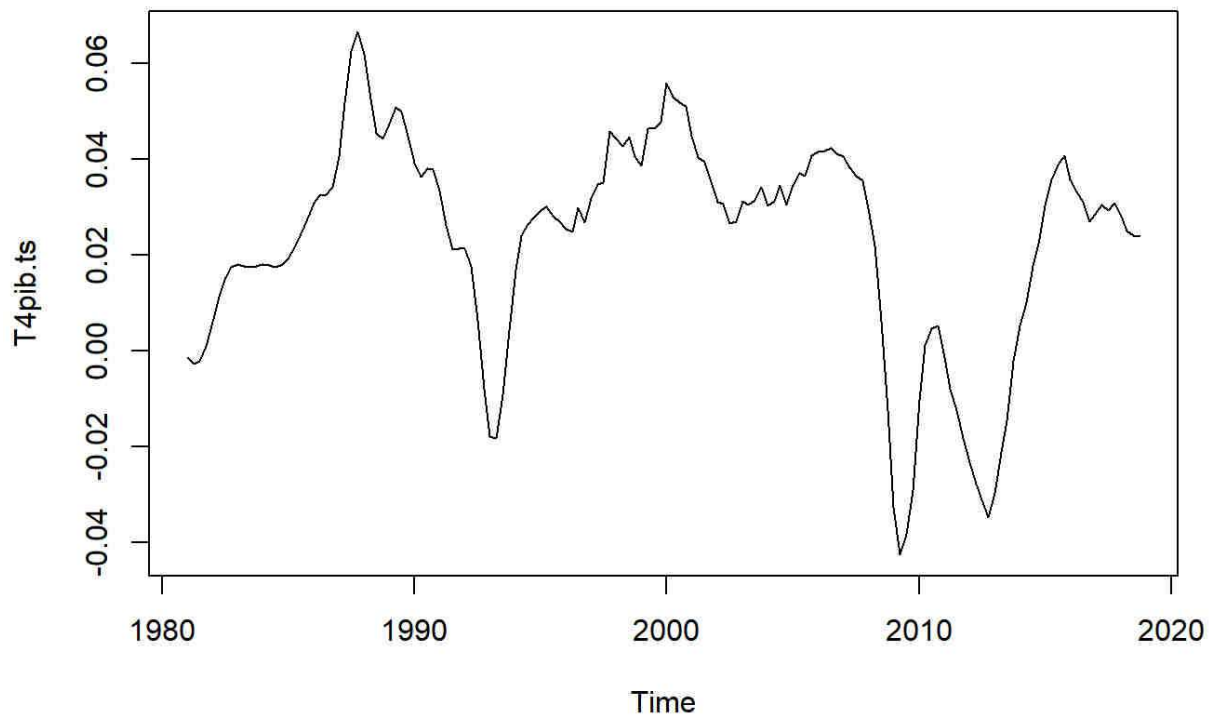
### 2.1 Leo los datos del PIB

```
# 1. Leo los datos del PIB (aprovecho los datos de la ley de okun ----
```

Leo los datos del PIB (aprovecho los datos de la Ley de Okun)

```
load("okundat.Rdata")
```

```
T4pib.ts<-window(okun[,5],start=1981, end=2018.75)  
plot(T4pib.ts)
```



## 2.2 Carga de datos de indicadores Adelantados: Indicadores de Confianza del Consumidor

```
# 2. Carga de datos de indicadores Adelantados: Indicadores de Confianza ----
```

Para cargar los datos de Indicador de Confianza del Consimidor **ICC** tenemos dos opciones **###** Descarga del **ICC** desde el Banco de España

```
# Sintesis de indicadores (Banco de Espa?a)  
si_1_1 <- read.csv("https://www.bde.es/webbde/es/estadis/infoest/series/si_1_1.csv", stringsAsFactor  
  s = FALSE, na.strings = "_")  
class(si_1_1)
```

```
## [1] "data.frame"
```



```
# Fila 2 - descripción de las variables
#str(si_1_1)
#si11.desc <- si_1_1[2,]
#si11.desc

# Datos ICC-índice de Confianza del Consumidor
ICC <- si_1_1[,c(1,3)]
ICC <- si_1_1[297:nrow(ICC),c(1,3)] # primer dato de la serie comienza en junio de 1986
head(ICC)
```

```
##           X SI_1_1.2
## 297 JUN 1986    -11.8
## 298 JUL 1986    -14.4
## 299 AGO 1986    -14.2
## 300 SEP 1986    -13.4
## 301 OCT 1986    -13.3
## 302 NOV 1986    -12.7
```

```
class(ICC)
```

```
## [1] "data.frame"
```

```
class(ICC[,2])
```

```
## [1] "character"
```

```
ICC_bde <- ts(as.numeric(ICC[,2]), start = c(1986,6), frequency=12)
```

## 2.2.1 Descarga del ICC desde el Eurostat

```
suppressMessages(library(eurostat))
suppressMessages(library(dplyr))
```

La librería eurostat permite descargar datos desde las librerías de tablas de eurostat

```
# Busco datos por tablas de Eurostat los indicadores de Confianza
#http://appsso.eurostat.ec.europa.eu/nui/show.do?dataset=ei_bssi_m_r2&lang=en#

query<-search_eurostat("Sentiment", type="all")
head(query,5)
```

```
## # A tibble: 1 x 8
##   title code type `last update of` `last table str` `data start`
##   <chr> <chr> <chr> <chr>          <chr>          <chr>
## 1 Sent~ ei_b~ data~ 30.10.2019      30.10.2019      1980M01
## # ... with 2 more variables: `data end` <chr>, values <chr>
```

```
temp.Eurostat<-get_eurostat("ei_bssi_m_r2",time_format = "num",type ="code", stringsAsFactors = FALSE,cache = FALSE) #,time_format = "raw",type ="code", stringsAsFactors = TRUE)
#Para ver que significan los campos
cc<-label_eurostat_vars(temp.Eurostat)
print(cc)
```

```
## [1] "Indicator"
## [2] "Seasonal adjustment"
## [3] "Geopolitical entity (reporting)"
## [4] "Period of time (a=annual, q=quarterly, m=monthly, d=daily, c=cumulated from January)"
```

```
# añadir los labels
temp.Eurostat<-label_eurostat(temp.Eurostat, code=names(temp.Eurostat))
head(temp.Eurostat %>%
      distinct(indic, indic_code),10)
```

```
## # A tibble: 6 x 2
##   indic                                indic_code
##   <chr>                                <chr>
## 1 Construction confidence indicator BS-CCI-BAL
## 2 Consumer confidence indicator    BS-CSMCI-BAL
## 3 Economic sentiment indicator     BS-ESI-I
## 4 Industrial confidence indicator   BS-ICI-BAL
## 5 Retail confidence indicator       BS-RCI-BAL
## 6 Services Confidence Indicator     BS-SCI-BAL
```

```
# Este es el significado de cada indicador
```

```
# BS-CCI-BAL: Construction confidence indicator
# BS-CSMCI-BAL: Consumer confidence indicator
# BS-ESI-I: Economic sentiment indicator
# BS-ICI-BAL: Industrial confidence indicator
# BS-RCI-BAL: Retail confidence indicator
# BS-SCI-BAL: Services Confidence Indicator
```

```
# Ahora tendría que seleccionar el dato de Españaa, Francia, Italia, Reino Unido y Alemania, por ejemplo
# Selecciono sólo el de España y Sólo el Indicador de Confianza del Consumidor
```

```
Sent.Eurostat<-temp.Eurostat %>%
  filter(geo_code %in% c("ES")) %>%
  filter(indic_code=="BS-CSMCI-BAL") %>%
  filter(s_adj_code=="NSA") %>%
  arrange(time_code)
```

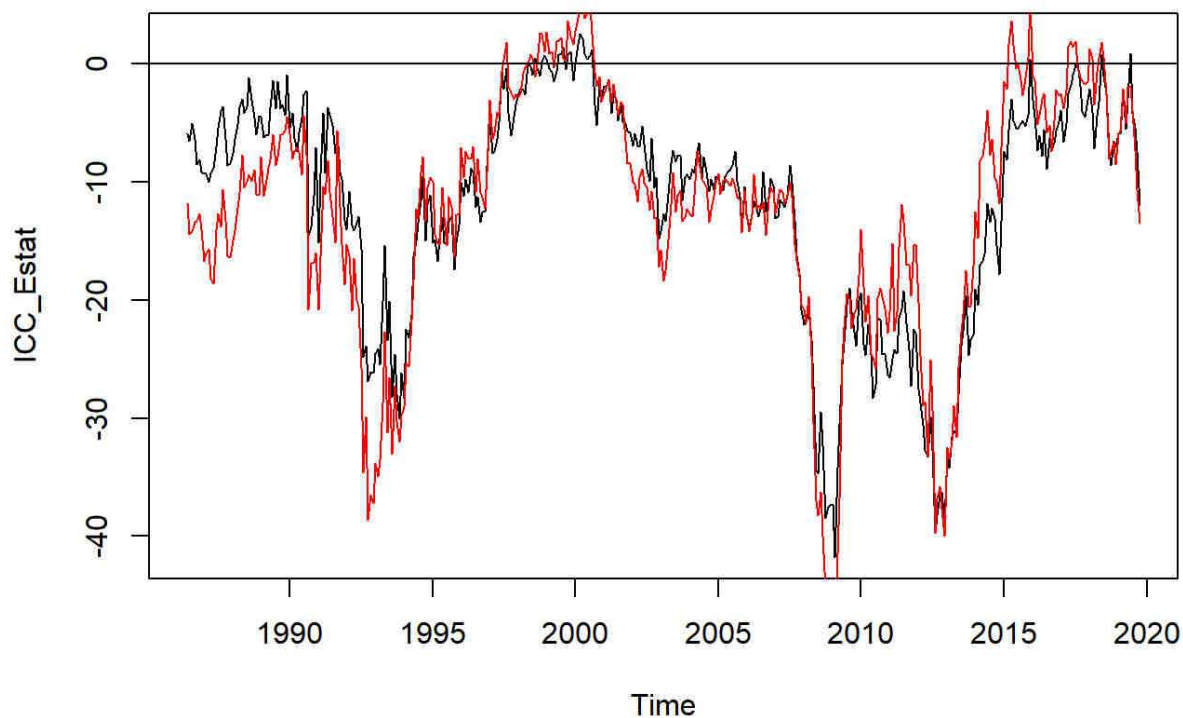
```
# Para comprobar automáticamente cuando comienza y cuando termina
startano<-floor(min(Sent.Eurostat$time_code))
starmes<-round((as.numeric(min(Sent.Eurostat$time_code))-startano)*12+1)
endano<-floor(max(Sent.Eurostat$time_code))
endmes<-round((as.numeric(max(Sent.Eurostat$time_code))-endano)*12+1)
endtrim<-round((endmes-1)*3/11+1)
```

```
ICC_Estat<-ts(as.numeric(Sent.Eurostat$values), start=c(startano,starmes), frequency = 12)
```

## 2.2.2 Dibujo las dos series ¿cómo interpretar el ICC?

aunque las dos series provienen de la misma fuente (la encuesta que realiza La comisión Europea) los datos de las dos fuentes no son idénticas, aunque proporcionan la misma información: ¿Cómo hay que interpretar los datos de la encuesta?

```
plot(ICC_Estat)
lines(ICC_bde,col=2)
abline(h=0)
```



selecciono los datos del Eurostat para seguir adelante con la práctica

```
ICC<-ICC_Estat
```

### 3 Transformo todas las series para que tengan la misma periodicidad (Trimestral)

```
# 3. Paso los datos mensuales a trimestrales ----
```

Las series PIB e ICC están en diferente frecuencia: ICC es mensual y PIB es trimestral Así que para poder correlacionar estas dos series tengo que **trimestralizar** las series mensuales

```
# Podemos utilizar la librería tempdisagg

suppressMessages(library (tempdisagg)) #ta agrega series, td trimestraliza/mensualiza series anuales

ICCT.ts1 <- tempdisagg::ta(ICC, conversion = "average", to="quarterly")
```

A veces esta librería tempdisagg da problemas (cuando no tiene datos de los tres meses de un trimestre no aprovecha la información mensual)

Así que también puedo trimestalizar manualmente

```
# Trimestralizo manualmente

ICCT<-ts.union(time(ICC),ICC,dframe=TRUE) %>%
  mutate(ano=floor(time(ICC)))%>%
  mutate(trim=0) %>%
  mutate(trim=ifelse(cycle(ICC) %in% 1:3,1,trim)) %>%
  mutate(trim=ifelse(cycle(ICC) %in% 4:6,2,trim)) %>%
  mutate(trim=ifelse(cycle(ICC) %in% 7:9,3,trim)) %>%
  mutate(trim=ifelse(cycle(ICC) %in% 10:12,4,trim))

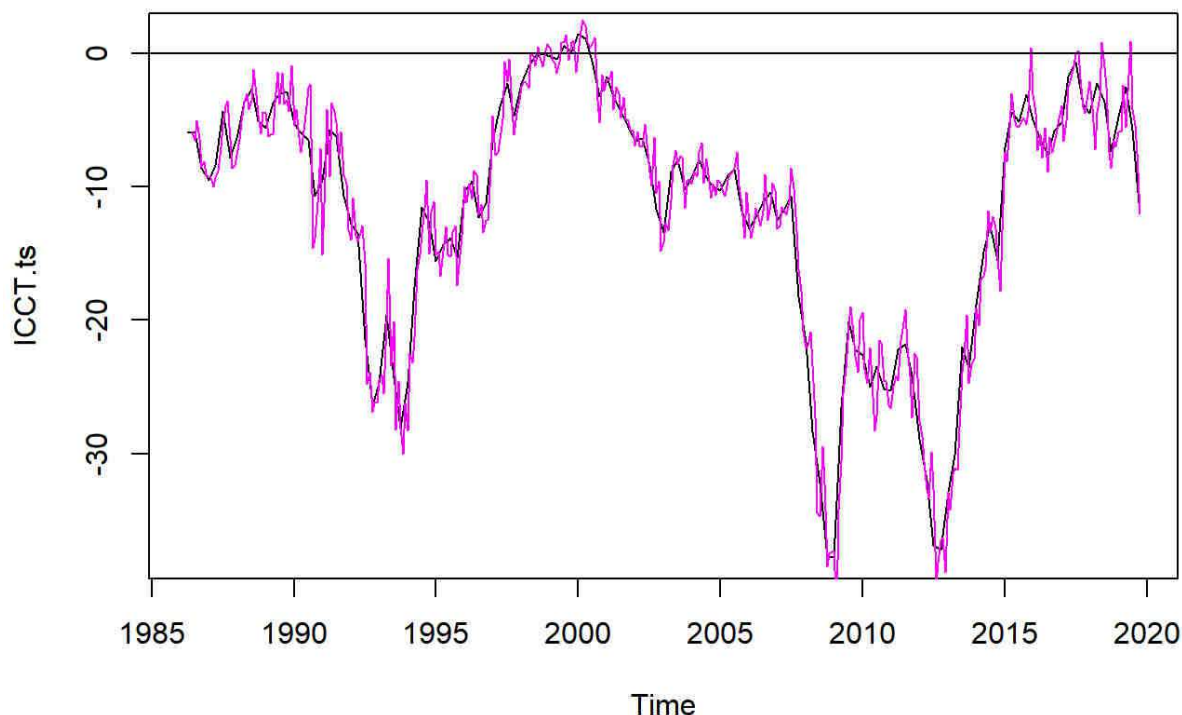
ICCT<-ICCT %>%
  group_by(ano,trim) %>%
  summarise(ICCTrim=mean(ICC,na.rm=TRUE))

ICCT.ts<-ts(ICCT$ICCTrim,start = c(1986,2), frequency = 4)

tail(ts.union(ICCT.ts,ICCT.ts1))
```

```
##           ICCT.ts  ICCT.ts1
## 2018 Q3  -3.633333 -3.633333
## 2018 Q4  -7.366667 -7.366667
## 2019 Q1  -4.733333 -4.733333
## 2019 Q2  -2.500000 -2.500000
## 2019 Q3  -5.700000 -5.700000
## 2019 Q4 -12.000000      NA
```

```
# Con la Trimestralización se ha suavizado la serie
plot(ICCT.ts)
lines(ICC, col=22)
abline(h=0)
```



## 4 Modelo Inicial: Asumiendo Correlaciones contemporaneas

Vamos a estimar un modelo de regresion en el que el PIB se explica por el ICC, esto es, vamos a utilizar los datos del ICC mensual (trimestralizado) para ayudar a predecir el PIB de un mismo trimestre.

```
# 4. MODELO inicial: correlaciones contemporaneas ----
```

```
suppressMessages(library(forecast))
```

```
# _____ #
```

### 4.1 Representación gráfica de las Dos series

Dibujo ambas series estandarizadas (para que estén en la misma escala)

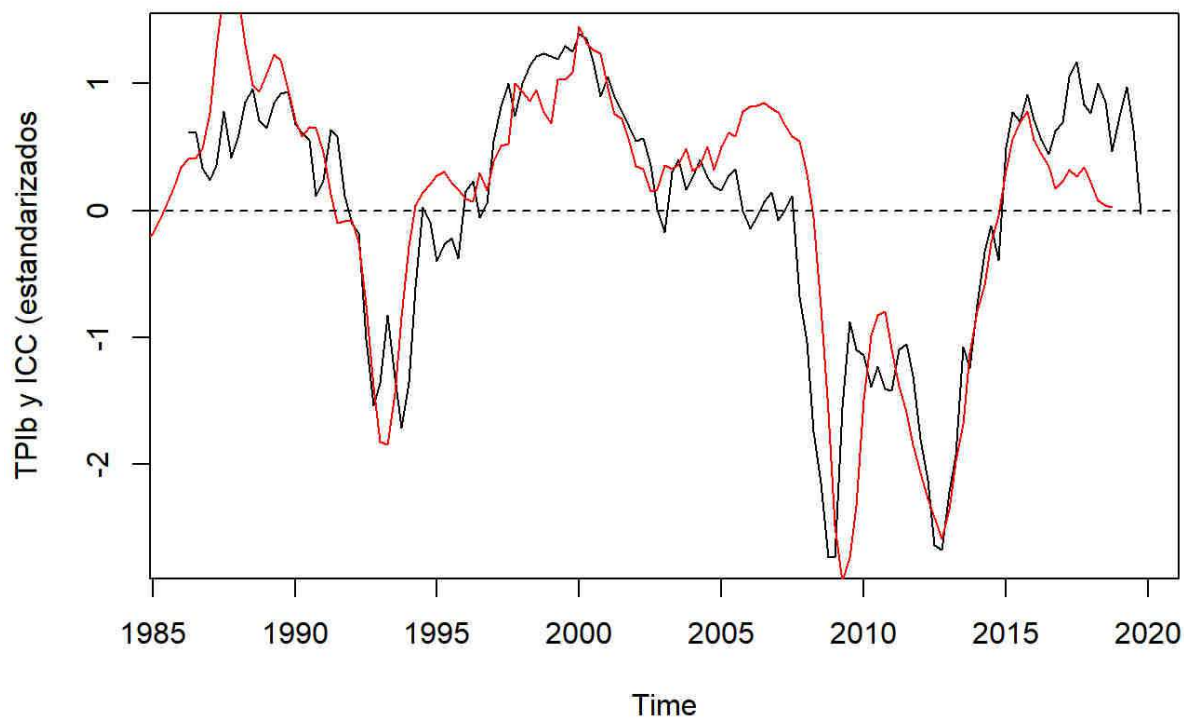
¿Cómo se mueven estas series a lo largo del tiempo?

```
# Gráfico de las dos series
```

```
plot((ICCT.ts-mean(ICCT.ts,na.rm = TRUE))/sd(ICCT.ts,na.rm = TRUE), type="l", ylab="TPIb y ICC (estandarizados)")
```

```
lines((T4pib.ts-mean(T4pib.ts,na.rm = TRUE))/sd(T4pib.ts,na.rm = TRUE), type="l", col=2)
```

```
abline(h=0, lty=2)
```



¿Están correlacionadas las dos series?

## 4.2 Análisis de la Estacionariedad en media y Cointegración

OJO: Como ambas son series temporales y vamos a hacer análisis de regresión nos interesa saber si series son **ESTACIONARIAS en media** (Sabrías como contrastar esta hipótesis?) y en su caso si están cointegradas

```
#
# Análisis de estacionariedad
ndiffs(T4pib.ts)
```

```
## [1] 1
```

```
ndiffs(ICCT.ts)
```

```
## [1] 1
```

```
# Parece que las series no son estacionarias (ambas son I(1)), para saber y la correlación es de ver-  
dad o es espúrea  
# Sabemos que tenemos que analizar la Cointegración  
  
# Análisis de cointegración  
suppressMessages(library(tseries))  
  
# construyo un dataframe con las dos series  
series<-ts.intersect(T4pib.ts,ICCT.ts, dframe = TRUE)  
  
# Test Cointegración H0: NO Cointegration  
# Phillips-Ouliaris Cointegration Test  
po.test(series,demean = TRUE, lshort=TRUE) #Phillips-Ouliaris Cointegration Test H0: NO Cointegrati-  
on
```

```
## Warning in po.test(series, demean = TRUE, lshort = TRUE): p-value smaller  
## than printed p-value
```

```
##  
## Phillips-Ouliaris Cointegration Test  
##  
## data: series  
## Phillips-Ouliaris demeaned = -33.996, Truncation lag parameter =  
## 1, p-value = 0.01
```

```
# otra librería para estimar el test Phillips-Ouliaris  
suppressMessages(library(urca))  
  
capocointest<-ca.po(series,demean=c("constant"),lag=c("short"))  
summary(capocointest)
```

```
##
## #####
## # Phillips and Ouliaris Unit Root Test #
## #####
##
## Test of type Pu
## detrending of series with constant only
##
##
## Call:
## lm(formula = z[, 1] ~ z[, -1])
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.045746 -0.005812 -0.001065  0.006655  0.033284
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  0.0496437   0.0017545    28.3   <2e-16 ***
## z[, -1]      0.0020978   0.0001146    18.3   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.01254 on 129 degrees of freedom
## Multiple R-squared:  0.7219, Adjusted R-squared:  0.7197
## F-statistic: 334.8 on 1 and 129 DF,  p-value: < 2.2e-16
##
##
## Value of test-statistic is: 50.686
##
## Critical values of Pu are:
##              10pct   5pct   1pct
## critical values 27.8536 33.713 48.0021
```

```
# si el estadístico es mayor que en valor en tablas
# o el valor crítico, entonces rechazamos la Ho nula
```

Están las series cointegradas ¿podemos seguir trabajando con estas series o tenemos que diferenciarlas?

## 4.3 Estimo el modelo lineal MCO

inicialmente podemos estimar un modelo lineal genral que nos sirva para predecir el PIB trimestral con los datos que vayan publicándose mensualmente del ICC

```
#series<-ts.intersect(T4pib.ts,ICCT.ts, dframe = TRUE)

modelo.lm<-lm(T4pib.ts ~ ICCT.ts, data = series, na.action=NULL)
#Para luego hacer comparaciones voy guardando los Criterios de información
modelo.lm$AIC<-AIC(modelo.lm)
modelo.lm$BIC<-BIC(modelo.lm)

stargazer(modelo.lm, type="text", keep = c("all","aic","bic"))
```

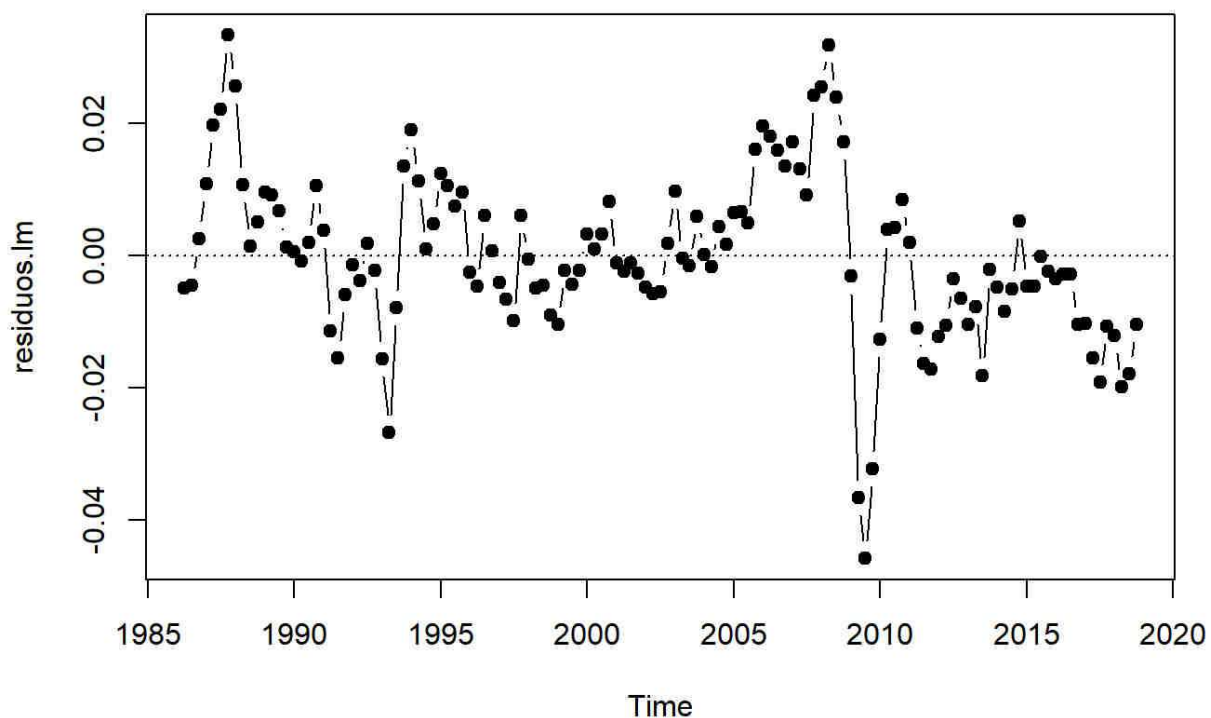


```
##
## =====
##                      Dependent variable:
##                      -----
##                      T4pib.ts
## =====
## Observations          131
## R2                    0.722
## Adjusted R2           0.720
## Akaike Inf. Crit.     -771.542
## Bayesian Inf. Crit.   -762.916
## Residual Std. Error    0.013 (df = 129)
## F Statistic            334.822*** (df = 1; 129)
## =====
## Note:                  *p<0.1; **p<0.05; ***p<0.01
```

Como estoy con series temporales es posible que tenga un problema de..... **Autocorrelación**

```
# Diagnósis de los residuos para detectar Autocorrelación
residuos.lm<-residuals(modelo.lm)

plot(residuos.lm, type="b", pch=19)
abline(h=0, lty=3)
```



```
# Tests de Autocorrelación
suppressMessages(library(lmtest))

# Durbin-Watson Test
lmtest::dwtest(modelo.lm) #H0: ausencia de autocorrelación de orden 1
```

```
##
## Durbin-Watson test
##
## data: modelo.lm
## DW = 0.41031, p-value < 2.2e-16
## alternative hypothesis: true autocorrelation is greater than 0
```

```
# Breusch-Godfrey Test
lmtest::bgtest(modelo.lm, order = 1) #H0: ausencia de autocorrelaci?n de orden p
```

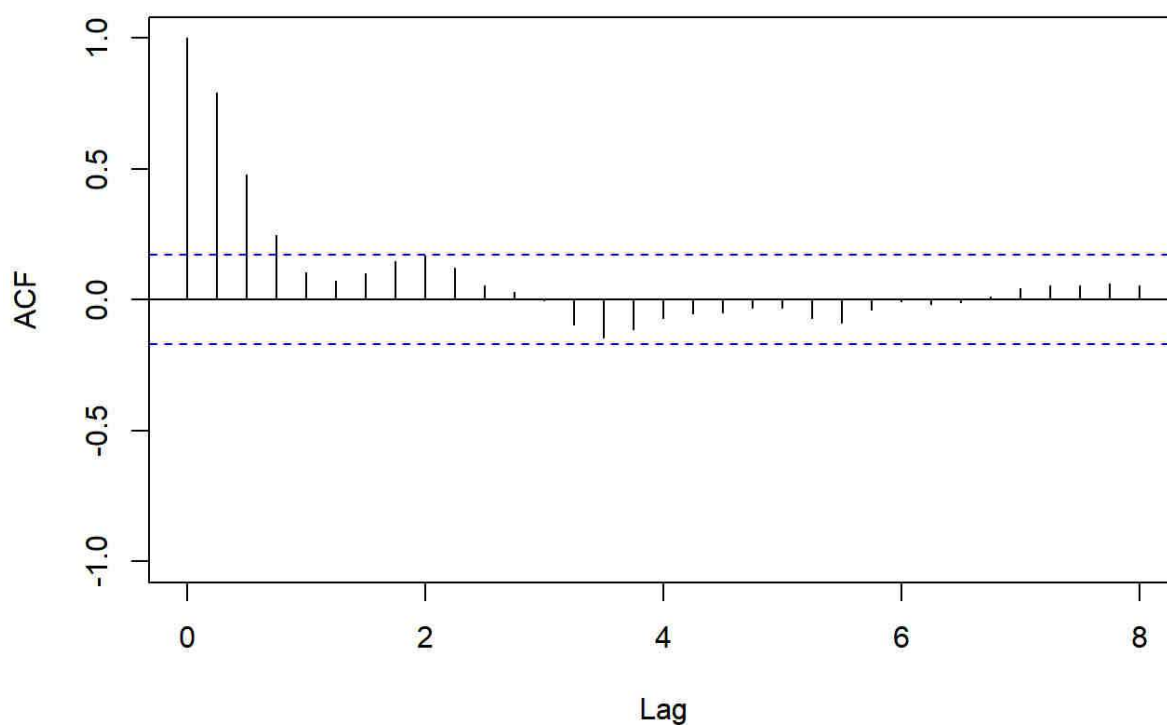
```
##
## Breusch-Godfrey test for serial correlation of order up to 1
##
## data: modelo.lm
## LM test = 82.572, df = 1, p-value < 2.2e-16
```

```
lmtest::bgtest(modelo.lm, order = 2)
```

```
##
## Breusch-Godfrey test for serial correlation of order up to 2
##
## data: modelo.lm
## LM test = 89.953, df = 2, p-value < 2.2e-16
```

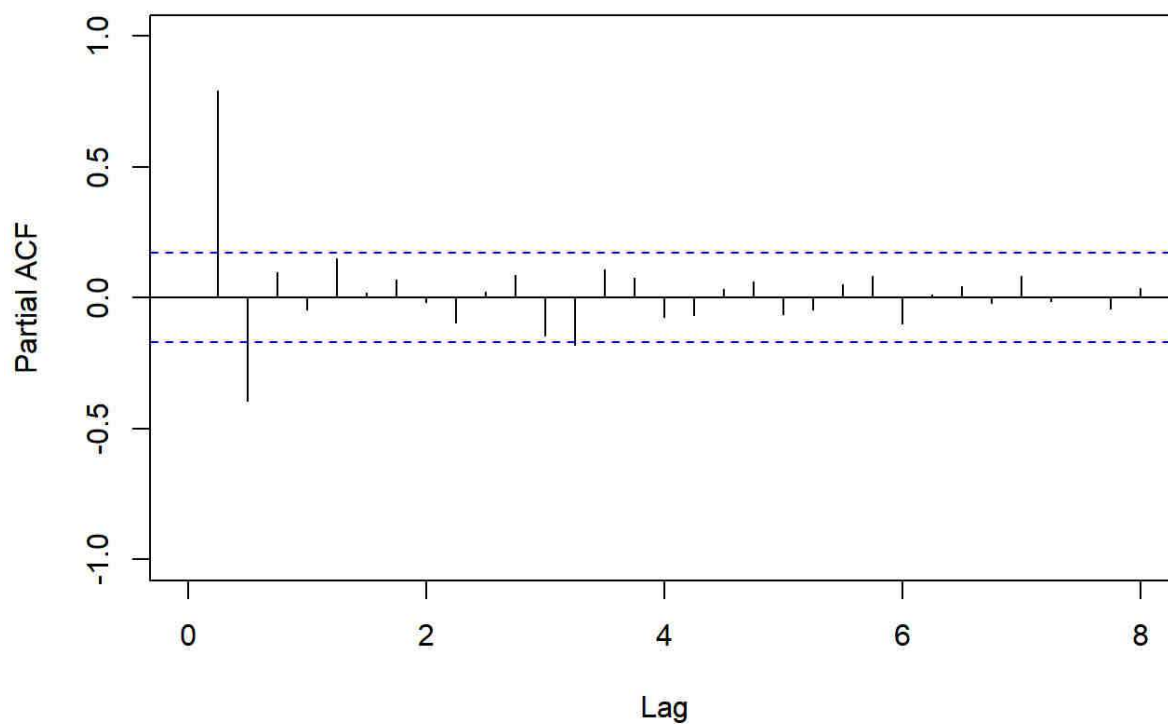
```
# correlograma
acf(residuos.lm, lag.max = 32, xlim=c(0,32/4), ylim=c(-1,1))
```

### Series residuos.lm



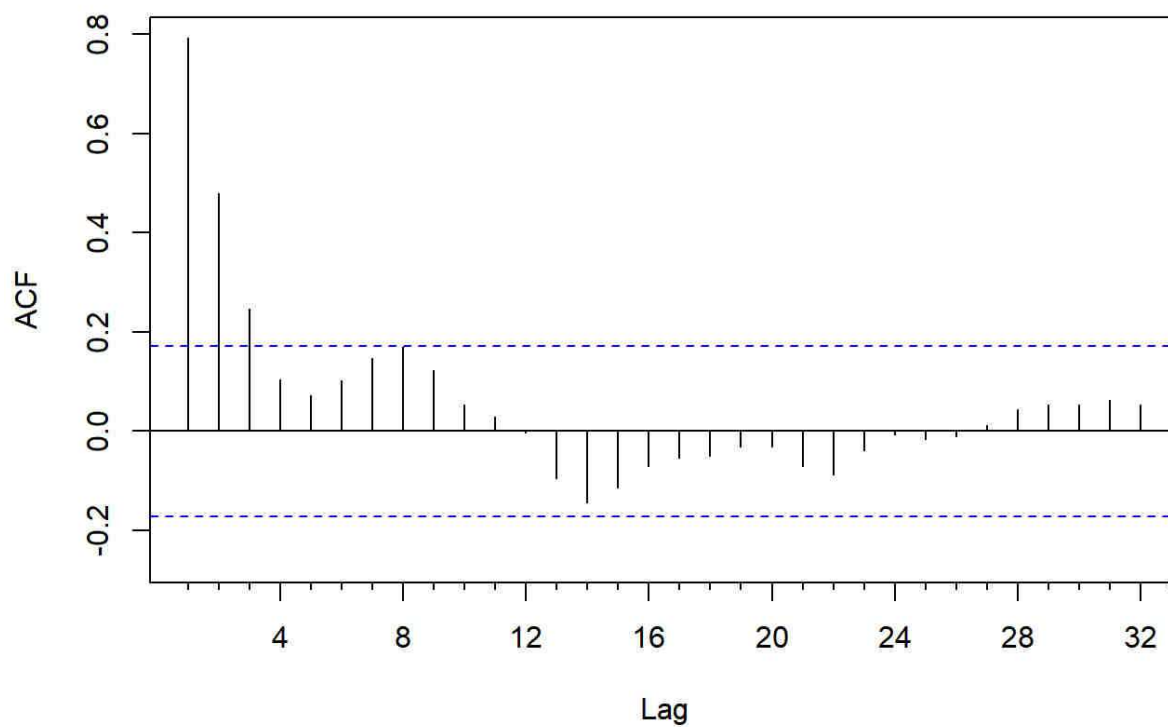
```
pacf(residuos.lm, lag.max = 32, xlim=c(0,32/4), ylim=c(-1,1))
```

### Series residuos.lm



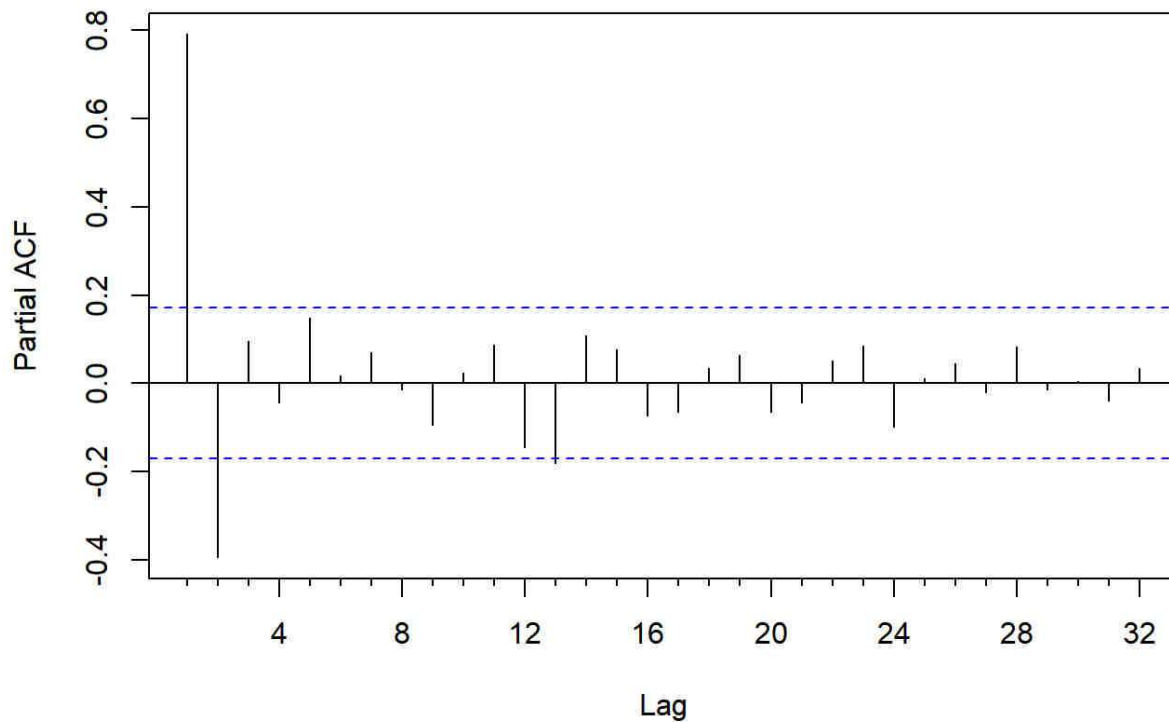
```
Acf(residuos.lm, lag.max = 32)
```

### Series residuos.lm



```
Pacf(residuos.lm, lag.max = 32)
```

## Series residuos.lm



```
# Ljung-Box test
for (i in 1:4){
  print(Box.test(residuos.lm, lag = i, type = c("Ljung-Box")))
}
```

```
##
## Box-Ljung test
##
## data:  residuos.lm
## X-squared = 83.987, df = 1, p-value < 2.2e-16
##
##
## Box-Ljung test
##
## data:  residuos.lm
## X-squared = 115, df = 2, p-value < 2.2e-16
##
##
## Box-Ljung test
##
## data:  residuos.lm
## X-squared = 123.3, df = 3, p-value < 2.2e-16
##
##
## Box-Ljung test
##
## data:  residuos.lm
## X-squared = 124.79, df = 4, p-value < 2.2e-16
```

Parece que tenemos un problema de autocorrelación en los residuos, así que sigo utilizando el estimador MCO, pero entonces tendré que estimar consistentemente la matriz de Var-Cov de los parámetros MCO

```
suppressMessages(library(sandwich))

modelo.lm.NW<-lmtest::coeftest(modelo.lm, vcov = sandwich::NeweyWest(modelo.lm))

stargazer(modelo.lm, modelo.lm.NW, type="text",keep.stat = c("n","rsq","ser","f","all","aic","bic"))
```

```
##
## =====
##                               Dependent variable:
##                               -----
##                               T4pib.ts
##                               OLS                coefficient
##                               test
##                               (1)                (2)
## -----
## ICCT.ts                0.002***                0.002***
##                               (0.0001)                (0.0003)
##
## Constant                0.050***                0.050***
##                               (0.002)                (0.004)
##
## -----
## Observations                131
## R2                0.722
## Akaike Inf. Crit.                -771.542
## Bayesian Inf. Crit.                -762.916
## Residual Std. Error    0.013 (df = 129)
## F Statistic    334.822*** (df = 1; 129)
## =====
## Note:                *p<0.1; **p<0.05; ***p<0.01
```

```
# _____ #
```

## 5 Correlación NO contemporanea e indicadores adelantados

Ya hemos comprobado que existe correlación entre ambas series temporales, ¿pero es dicha correlación contemporanea?

Para responder a esta pregunta analizaremos - Coeficiente de Correlación cruzada - Test de Granger de Causalidad

## Coeficiente de Correlación Cruzada

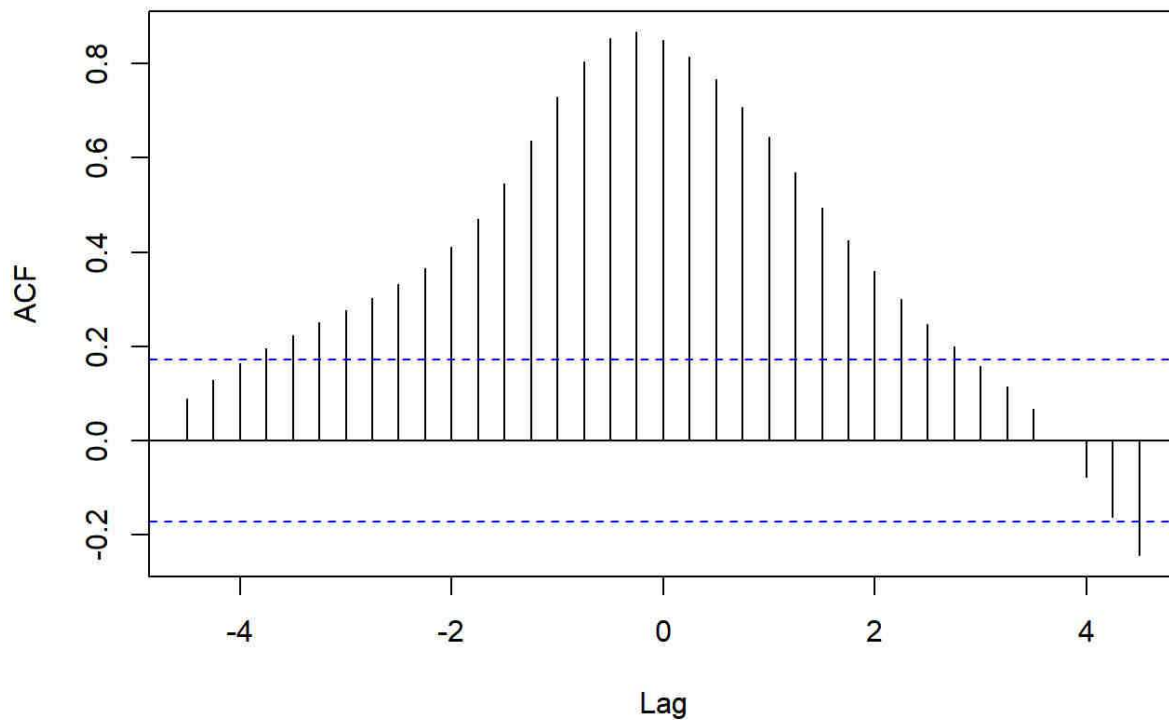
```
# 5. Correlación NO contemporanea e indicadores adelantados ----

#Ccf: Coeficiente de correlacion cruzada [xt+k, yt], Ccf(x,y)

# Para k=0 (retardo =0) tenemos la correlación contemporánea,
# dónde se alcanza la correlación máxima entre ambas series (en qué retardo)

ccf(ICCT.ts, T4pib.ts)
print(ccf(ICCT.ts, T4pib.ts))
```

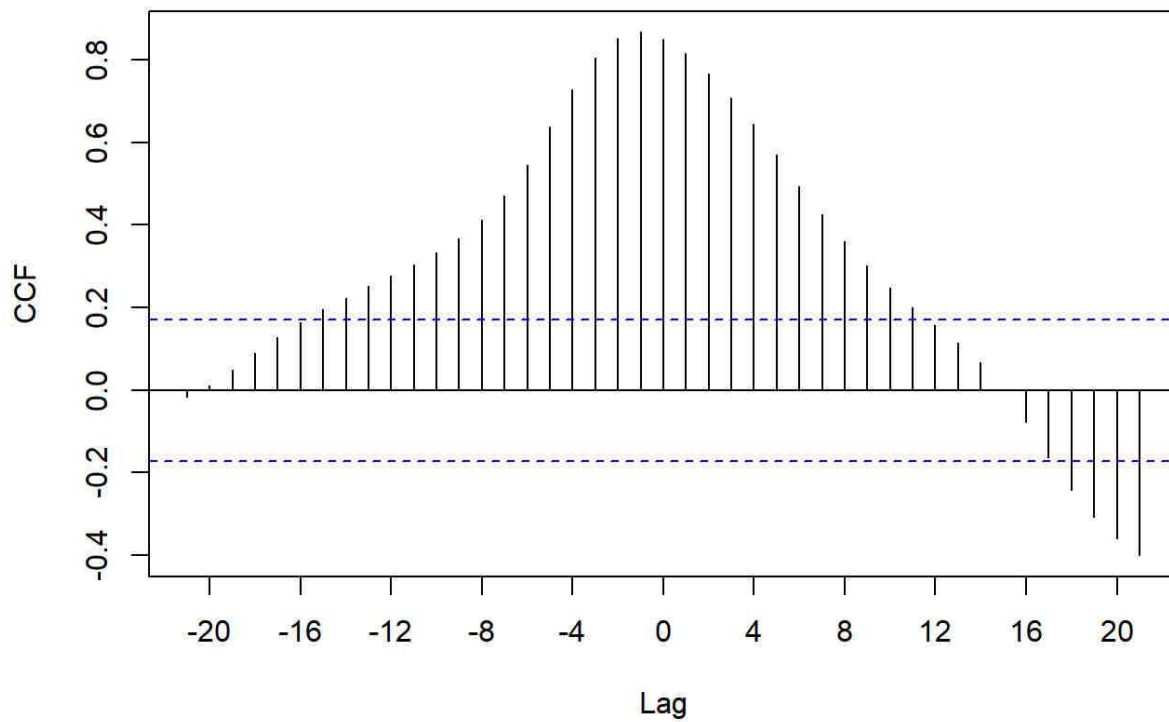
## ICCT.ts & T4pib.ts



```
##
## Autocorrelations of series 'X', by lag
##
## -4.50 -4.25 -4.00 -3.75 -3.50 -3.25 -3.00 -2.75 -2.50 -2.25
## 0.089 0.128 0.164 0.195 0.223 0.251 0.276 0.303 0.332 0.366
## -2.00 -1.75 -1.50 -1.25 -1.00 -0.75 -0.50 -0.25 0.00 0.25
## 0.411 0.470 0.546 0.636 0.729 0.805 0.853 0.867 0.850 0.815
## 0.50 0.75 1.00 1.25 1.50 1.75 2.00 2.25 2.50 2.75
## 0.766 0.708 0.644 0.570 0.494 0.425 0.360 0.301 0.247 0.200
## 3.00 3.25 3.50 3.75 4.00 4.25 4.50
## 0.157 0.115 0.066 0.001 -0.077 -0.162 -0.243
```

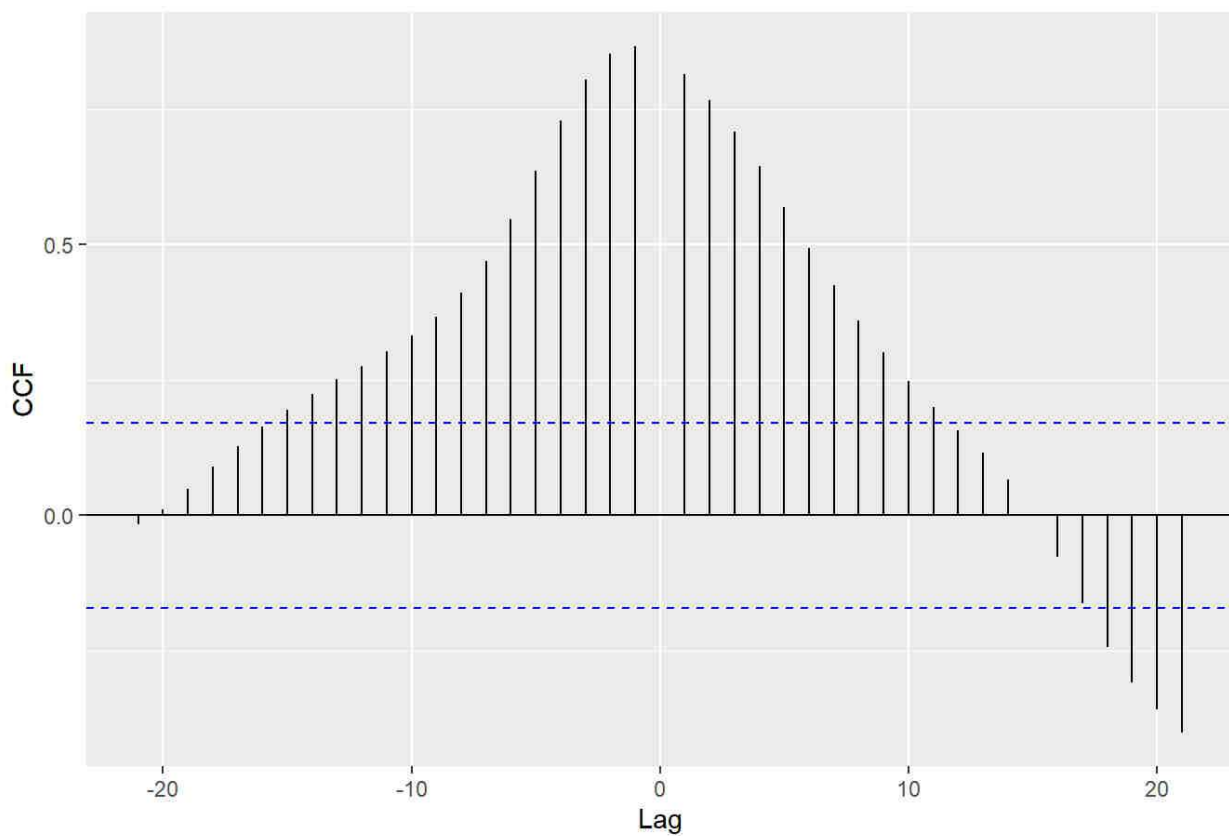
```
# o esta función
Ccf(ICCT.ts, T4pib.ts)
```

## ICCT.ts & T4pib.ts



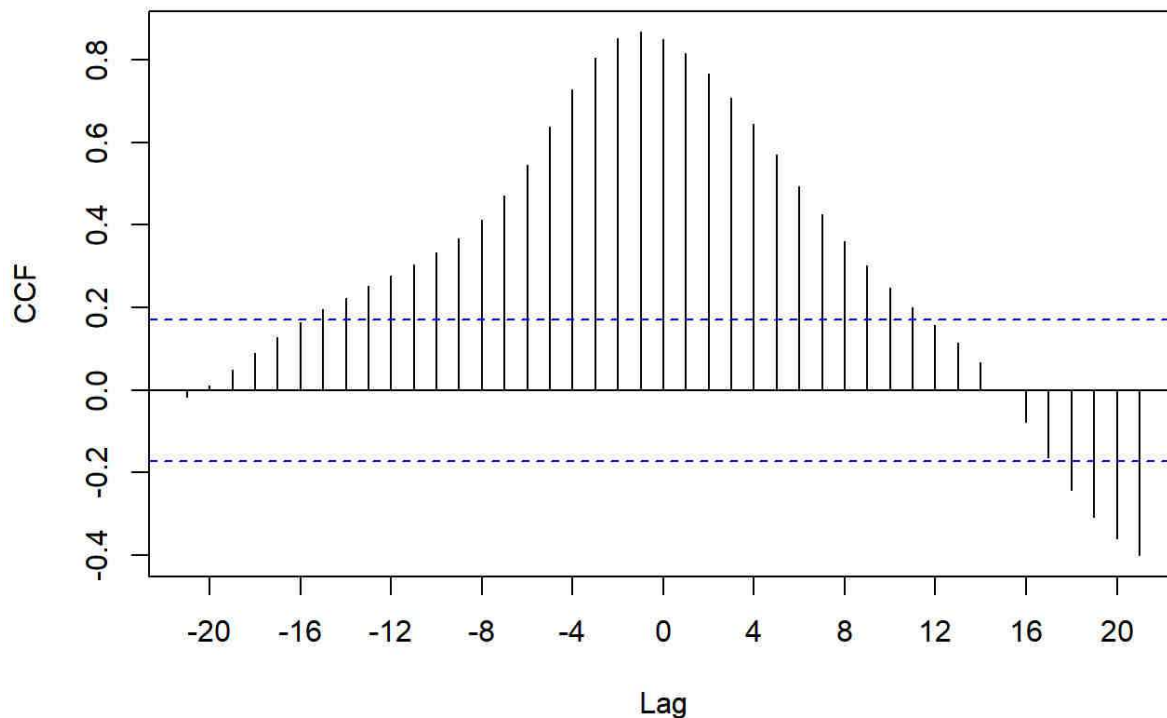
```
# o esta  
ggCcf(ICCT.ts, T4pib.ts)
```

Series: ICCT.ts & T4pib.ts



```
print(Ccf(ICCT.ts,T4pib.ts))
```

## ICCT.ts & T4pib.ts



```
##
## Autocorrelations of series 'X', by lag
##
##   -21   -20   -19   -18   -17   -16   -15   -14   -13   -12
## -0.017  0.010  0.048  0.089  0.128  0.164  0.195  0.223  0.251  0.276
##   -11   -10    -9    -8    -7    -6    -5    -4    -3    -2
##  0.303  0.332  0.366  0.411  0.470  0.546  0.636  0.729  0.805  0.853
##    -1     0     1     2     3     4     5     6     7     8
##  0.867  0.850  0.815  0.766  0.708  0.644  0.570  0.494  0.425  0.360
##     9     10     11     12     13     14     15     16     17     18
##  0.301  0.247  0.200  0.157  0.115  0.066  0.001 -0.077 -0.162 -0.243
##    19     20     21
## -0.308 -0.359 -0.400
```

Imprime los valores para todos los retardos (dibujados arriba)

¿dónde se da el máximo valor muy próximo de los dos primeros retardos (no para lag=0 - no es corr. contemporánea)?

Es la máxima correlación contemporánea? ¿o existe alguna de las dos series que adelante a la otra?

# \_\_\_\_\_ #

## 5.1 Test de causalidad de Granger

Este test permite contrastar si existe algún **indicador adelantado**, esto es, conocer si el valor actual de una serie sirve para adelantar los valores futuros de la otra, o si por el contrario la series son **contemporáneas**

Notese que si el ICC fuese un indicador adelantado podría utilizarse no sólo para predecir el valor actual del PIB aprovechando que tiene periodicidad mensual, también podría utilizarse para saber que hará el PIB en los trimestres siguientes



```
# Test de causalidad de Granger
suppressMessages(library(lmtest))

# Sentido 1: X => Y
# H0: X (ICC) NO CAUSA (en sentido de Granger) a Y (PIB)
# H0: beta1=beta2=0
# Modelo 1: Y(t) | cte alpha1*Y(t-1) alpha2*Y(t-2) beta1*X(t-1) beta2*X(t-2)
# Modelo 2: Y(t) | cte alpha1*Y(t-1) alpha*Y(t-2)
grangertest(ICCT.ts, T4pib.ts, order=3)
```

```
## Granger causality test
##
## Model 1: T4pib.ts ~ Lags(T4pib.ts, 1:3) + Lags(ICCT.ts, 1:3)
## Model 2: T4pib.ts ~ Lags(T4pib.ts, 1:3)
##   Res.Df Df       F    Pr(>F)
## 1      121
## 2      124 -3 10.819 2.385e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
# Sentido 2: Y => X
# H0: Y (PIB) NO CAUSA (en sentido de Granger) a X (ICC)
# H0: beta1=beta2=0
# Modelo 1: X(t) | cte alpha1*X(t-1) alpha2*X(t-2) beta1*Y(t-1) beta2*Y(t-2)
# Modelo 2: X(t) | cte alpha1*X(t-1) alpha*X(t-2)

grangertest(T4pib.ts, ICCT.ts, order=3)
```

```
## Granger causality test
##
## Model 1: ICCT.ts ~ Lags(ICCT.ts, 1:3) + Lags(T4pib.ts, 1:3)
## Model 2: ICCT.ts ~ Lags(ICCT.ts, 1:3)
##   Res.Df Df       F    Pr(>F)
## 1      121
## 2      124 -3 2.3699 0.07393 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
#_____#
```

## 6 Estimación del Modelo con indicador adelantado

Una vez que se ha detectado que existe un indicador adelantado, que el ICC adelanta el comportamiento del PIB Podemos intentar construir un modelo para intentar aprovechar esa información del indicador adelantado

```
# Estimación del modelo (teniendo en cuenta el sentido de la causalidad)
# PIB = b1 + b2*ICC(t-1) + error

# para estimar el modelo es necesario que las series tengan la misma longitud
#modelo <- T4pib.ts ~ lag(ICCT, k=1)
#modelo.lm <- lm(modelo)

# Retardo un periodo, pero podría ir eligiendo diferentes retardos
seriex<-stats::lag(ICCT.ts,k=-1)
modelo<- T4pib.ts ~ seriex

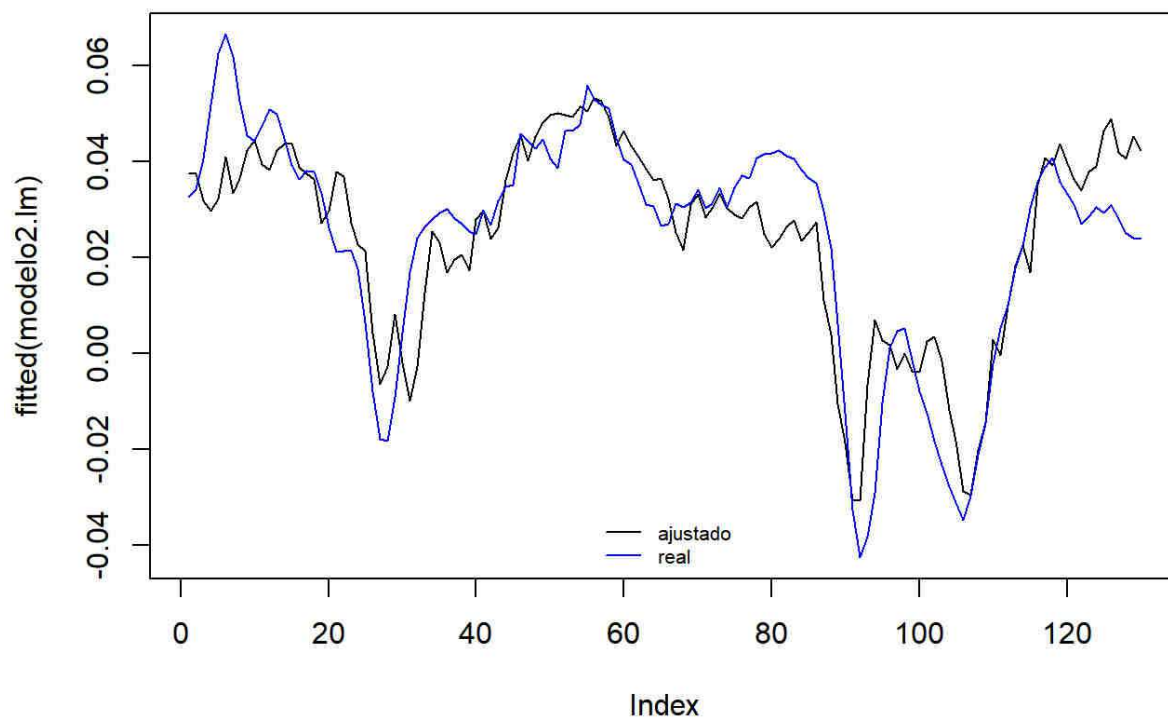
series<-ts.intersect(T4pib.ts,seriex)

modelo2.lm<-lm(modelo, data=series,na.action=NULL) #ojo, na.action=NULL para las series
salida.lm<-summary(modelo2.lm)
print(salida.lm)
```

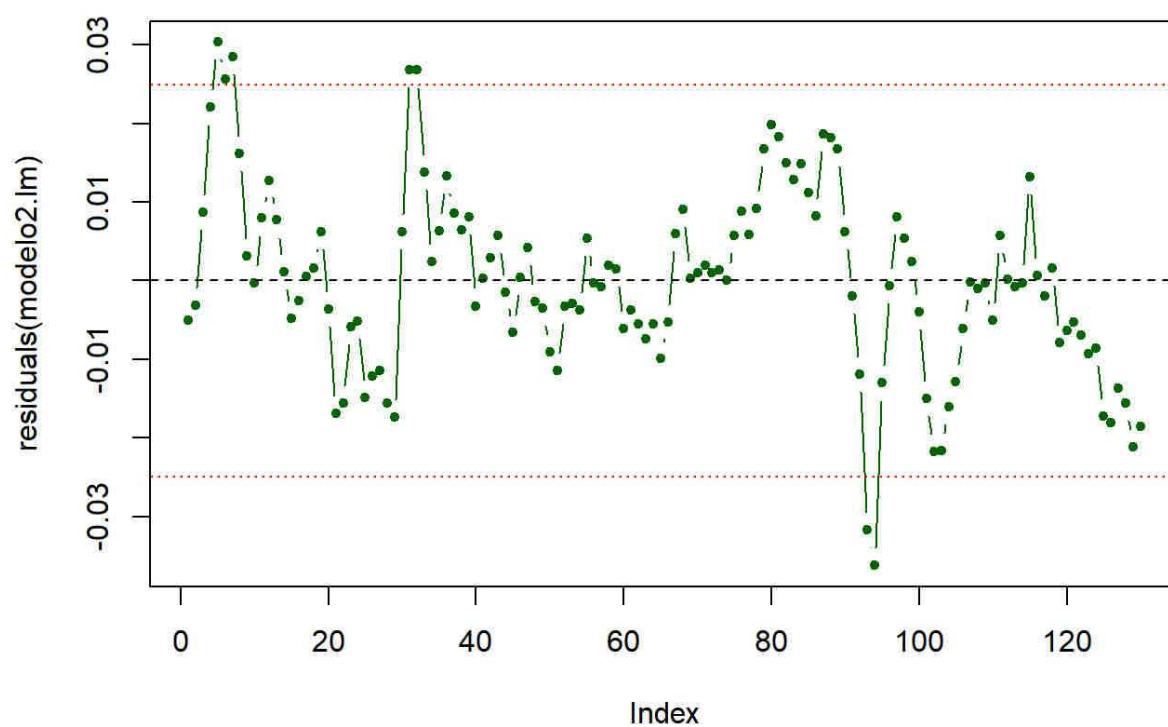
```
##
## Call:
## lm(formula = modelo, data = series, na.action = NULL)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.036210 -0.006075 -0.000214  0.006285  0.030325
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  0.0502069   0.0016653   30.15  <2e-16 ***
## seriex       0.0021436   0.0001085   19.76  <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.01186 on 128 degrees of freedom
## Multiple R-squared:  0.7531, Adjusted R-squared:  0.7511
## F-statistic: 390.3 on 1 and 128 DF, p-value: < 2.2e-16
```

```
modelo2.lm$AIC<-AIC(modelo2.lm)
modelo2.lm$BIC<-BIC(modelo2.lm)

# Diagnósis del modelo
# Gráfico de la serie y del valor ajustado
#límites para el gráfico para que no se corte
ylimitesdbj<-c(min(c(min(series[,1]),min(fitted(modelo2.lm)))),max(c(max(series[,1]),max(fitted(modelo2.lm)))))
plot(fitted(modelo2.lm), type="l", ylim= ylimitesdbj)
lines(as.numeric(series[,1]), type="l", col="blue")
legend("bottom",legend = c("ajustado", "real"),lty=c(1), cex=0.65, lwd=c(1,1),
      bty="n",
      col=c("black","blue"))
```

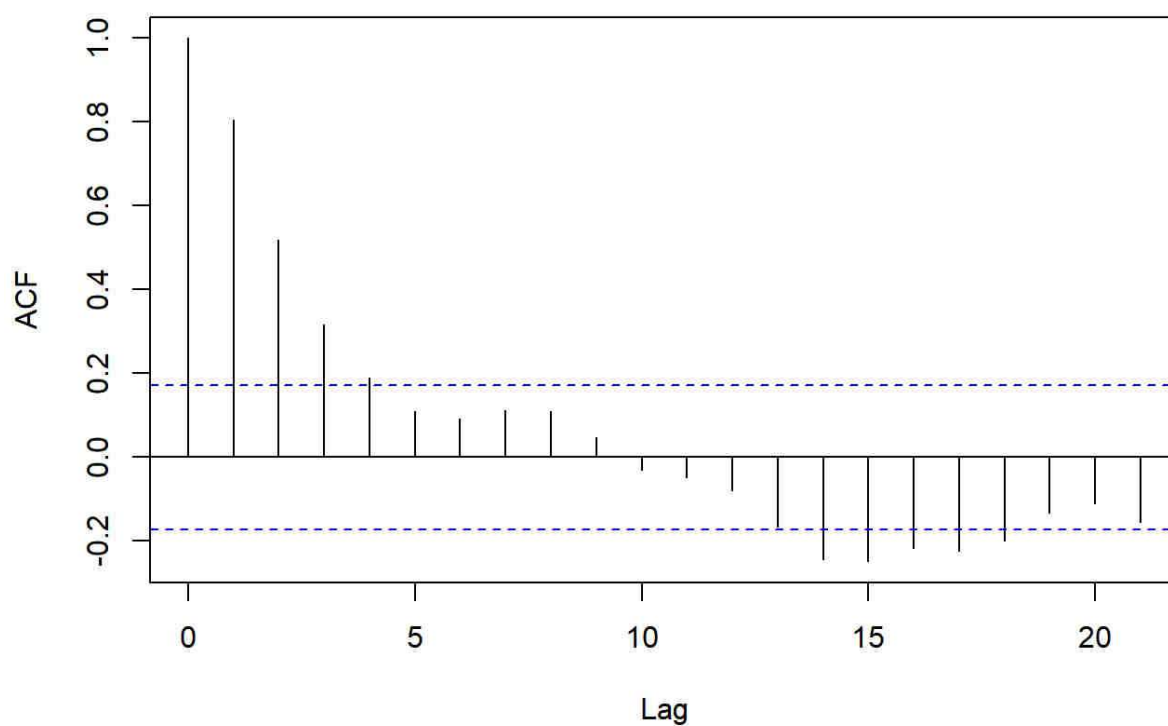


```
# Gráfico de los residuos
plot(residuals(modelo2.lm), type="b", col="darkgreen", pch=20)
abline(h=0, lty=2, col="grey2" )
abline(h=2*sd(residuos.lm), lty=3, col="red" )
abline(h=-2*sd(residuos.lm), lty=3, col="red" )
```



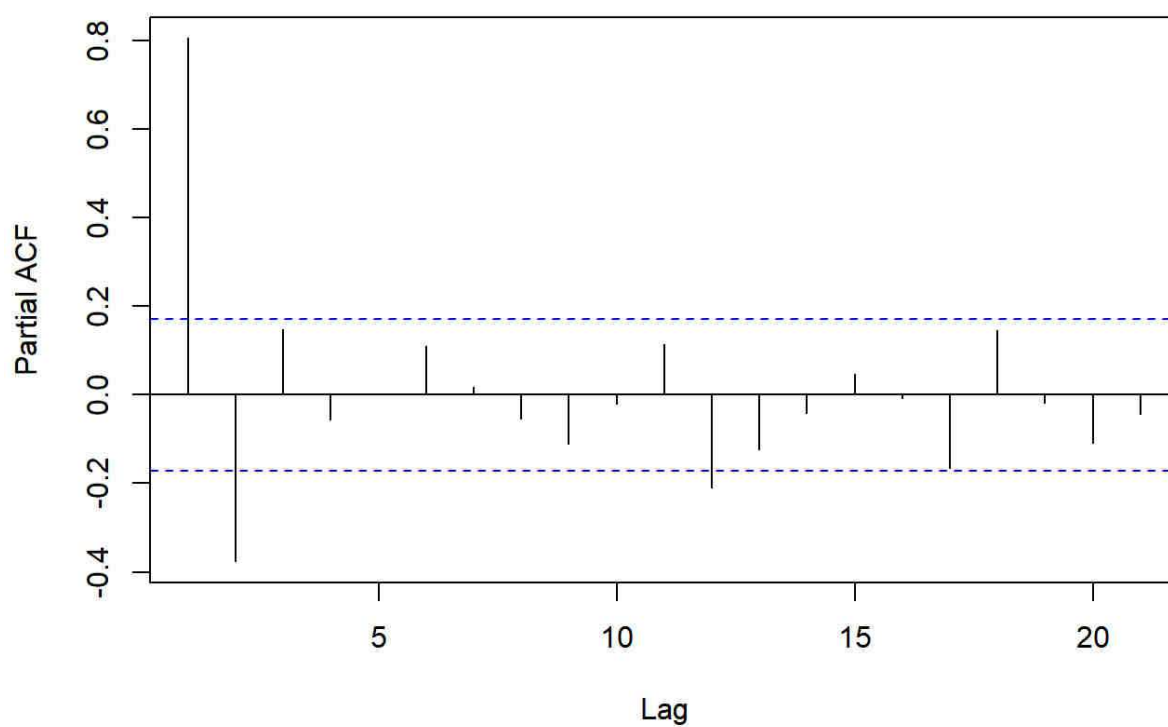
```
# detección de autocorrelación  
acf(residuals(modelo2.lm))
```

**Series residuals(modelo2.lm)**



```
pacf(residuals(modelo2.lm))
```

**Series residuals(modelo2.lm)**



```
# Como los residuos están autocorrelacionados estimo Consistentemente la matriz Var-Cov de los estimadores
modelo2.lm.NW<-lmtest::coefest(modelo2.lm, vcov = sandwich::NeweyWest(modelo2.lm))

stargazer(modelo.lm, modelo.lm.NW, modelo2.lm, modelo2.lm.NW , type="text")
```

```
##
## =====
##                               Dependent variable:
## -----
##               T4pib.ts               T4pib.ts
##               OLS               coefficient OLS               coefficient
##               test               test
##               (1)               (2)               (3)               (4)
## -----
## ICCT.ts               0.002***               0.002***
##               (0.0001)               (0.0003)
##
## seriex               0.002***               0.002***
##               (0.0001)               (0.0003)
##
## Constant               0.050***               0.050***
##               (0.002)               (0.004)               (0.002)               (0.005)
## -----
## Observations               131               130
## R2               0.722               0.753
## Adjusted R2               0.720               0.751
## Akaike Inf. Crit.               -771.542               -780.178
## Bayesian Inf. Crit.               -762.916               -771.575
## Residual Std. Error               0.013 (df = 129)               0.012 (df = 128)
## F Statistic               334.822*** (df = 1; 129)               390.335*** (df = 1; 128)
## =====
## Note:                               *p<0.1; **p<0.05; ***p<0.01
```

## 6.1 Predicción con los los indicadores adelantados

```
tail(series)
```

```
##           T4pib.ts      seriex
## 2017 Q3 0.02924071 -1.733333
## 2017 Q4 0.03089747 -0.600000
## 2018 Q1 0.02824484 -3.866667
## 2018 Q2 0.02505401 -4.466667
## 2018 Q3 0.02415217 -2.266667
## 2018 Q4 0.02383999 -3.633333
```

```
# El último dato del PIB termina el 4 trimestre de 2018, así que utilizo todos los datos
# disponibles del indicador adelantado para hacer predicciones
```

```
tail(seriex,10)
```

```
##           Qtr1      Qtr2      Qtr3      Qtr4
## 2017                -0.600000
## 2018 -3.866667 -4.466667 -2.266667 -3.633333
## 2019 -7.366667 -4.733333 -2.500000 -5.700000
## 2020 -12.000000
```

```
newseriex<-data.frame(seriex=seriex[time(seriex)>=2019])
```

```
# puedo obtener las predicciones con dos funciones
predict(modelo2.lm,newdata = newseriex, se.fit = TRUE, interval="prediction", level=0.95)
```

```
## $fit
##           fit           lwr           upr
## 1 0.03441570 0.0108462148 0.05798518
## 2 0.04006051 0.0164604658 0.06366056
## 3 0.04484789 0.0212113511 0.06848442
## 4 0.03798836 0.0144011076 0.06157562
## 5 0.02448368 0.0009351009 0.04803226
##
## $se.fit
##           1           2           3           4           5
## 0.001154451 0.001304219 0.001463294 0.001243723 0.001039814
##
## $df
## [1] 128
##
## $residual.scale
## [1] 0.0118557
```

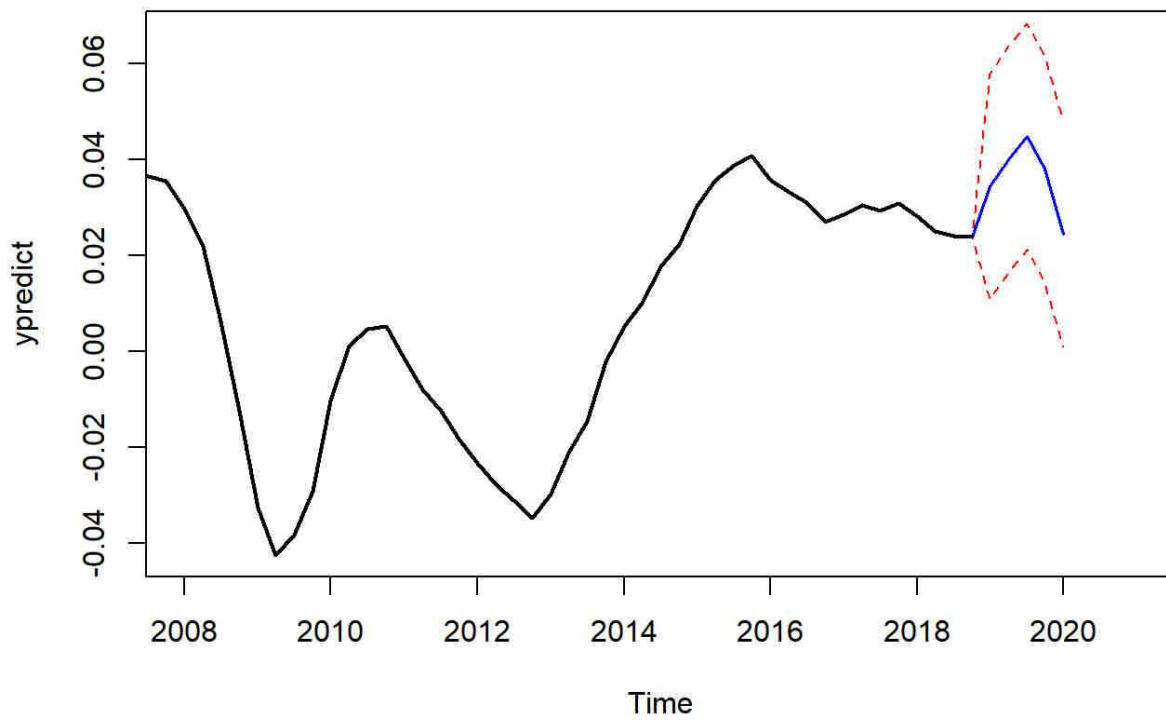
```
forecast(modelo2.lm,newdata = newseriex)
```

```
## Point Forecast      Lo 80      Hi 80      Lo 95      Hi 95
## 1 0.03441570 0.019070942 0.04976045 0.0108462148 0.05798518
## 2 0.04006051 0.024695859 0.05542516 0.0164604658 0.06366056
## 3 0.04484789 0.029459477 0.06023629 0.0212113511 0.06848442
## 4 0.03798836 0.022632038 0.05334469 0.0144011076 0.06157562
## 5 0.02448368 0.009152535 0.03981483 0.0009351009 0.04803226
```

```
prediccion<-forecast(modelo2.lm,newdata = newseriex)

ypredict<-ts(c(as.numeric(series[,1]),prediccion$mean),
             start=start(series[,1]), frequency=frequency(series[,1]))
ypredict_low<-ts(c(as.numeric(series[,1]),prediccion$lower[,2]),
                start=start(series[,1]), frequency=frequency(series[,1]))
ypredict_upp<-ts(c(as.numeric(series[,1]),prediccion$upper[,2]),
                start=start(series[,1]), frequency=frequency(series[,1]))

plot(ypredict, col=4, lwd=1.5, xlim=c(2008,2021))
lines(ypredict_low, col=2, lty=2, lwd=1)
lines(ypredict_upp, col=2, lty=2, lwd=1)
lines(series[,1], col=1, lwd=2)
```



# \_\_\_\_\_ #

**TAREA: Y SI EXISTIESN MAS INDICADORES ADELANTADO? ¿SE PODRIA UTILIZAR EL CONJUNTO DE TODOS ESOS INDICADORES PARA HACER LA PREDICCIÓN DEL PIB**



Proyecto de innovación  
“Innova Docencia” de la UCM  
Convocatoria 2018-2019  
Proyecto número 275

**R-adaptación de la asignatura de métodos econométricos en economía y finanzas del  
Grado de Estadística Aplicada**

Facultad de Estudios Estadísticos  
Departamento de Economía Aplicada, Pública y Política

**Título**

***Práctica 13: Análisis univariante de la serie del PIB y ARIMAX con tasas  
de ocupados***

**Correspondiente al**

**Capítulo 6 MODELOS CON DATOS DE SERIES TEMPORALES III: MODELOS  
DINAMICOS ARIMA y ARIMAX**

- Series temporales estacionarias y Orden de integración I(d)
- Modelos ARIMA
- Estimación, Diagnóstico de modelos ARIMA
- Extensión: Modelos ARIMAX

***Práctica 13: Análisis univariante de la serie del PIB y ARIMAX con tasas de  
ocupados***



# Modelos de regresión con series temporales estacionarias: Modelos dinámicos ARIMA y ARIMAX

## Asignatura Métodos Econométricos en Economía y Finanzas

Profesores  
Loren Escot, Alicia Pérez-Alonso y Julio E. Sandubete

Dpto Economía Aplicada, Pública y Política  
Facultad Estudios Estadísticos - Universidad Complutense de Madrid

Curso académico 2018/2019

### Abstract

En esta práctica se revisará la estimación de modelos dinámicos **ARMA** y modelos **ARMAX** cuando las series empleadas resultan estacionarias.

- 1 Librería necesaria
- 2 Carga de datos—
- 3 Modelos Univariantes ARIMA
  - 3.1 Estimación de un AR(0,0)
  - 3.2 Estimación de un AR(1,0)
- 4 Estimación modelo automático: auto.arima
- 5 PREDICCIONES
- 6 Modelos Multivariantes: ARMAX

```
#=====#
```

```
# Modelos de regresión Dinámicos----
```

## 1 Librería necesaria

Existen distintas librerías que incluyen funciones para estimar **modelos ARMA**. nosotros vamos a utilizar en esta ocasión la librería `forecast`

```
suppressMessages(library(forecast))
```

## 2 Carga de datos—

Vamos a utilizar las series de tasas de crecimiento y empleo para estimar modelos univariantes ARMA

```
load("okundat.Rdata")

T4pib.ts<-window(okun[,5], end=2018.75)
pib.ts<-window(okun[,2], end=2018.75)
```

# 3 Modelos Univariantes ARIMA

Los modelos ARIMA: son modelos dinámicos de regresión que suponen que los valores presentes de una variable dependen de los valores pasados de esa misma variable, más un error y más los errores del pasado

Son modelos que sólo pueden aplicarse sobre series estacionarias en media. Por eso, antes de comenzar la estimación de modelos ARMA hay que transformar la serie para conseguir que sea estacionaria tanto en varianza (quizas transformando la serie en logaritmos) como en media (diferenciando la serie tantas veces como sea necesario). De hecho, la "I" de los modelos ARIMA, hace referencia precisamente al orden de integración de la serie (número de veces que hay que diferenciar la serie)

Utilizaremos además **modelos automáticos**. Nos preguntaremos Qué orden autorregresivo AR(p), y de media móvil MA(q) proporciona el mejor ajuste a la serie temporal.

La selección del mejor modelo se basa en elegir aquél que proporcione mejores errores pasados corregidos por el número de parámetros utilizados, esto es, minimizando algún criterio de información (AIC, SIC, etc).

```
# 1.Modelos Univariantes ARIMA ----

# ARIMA: Modelo de regresión de una variable sobre los valores pasados
#       de sí mismo y sobre los errores del pasado

# Modelos automáticos: ¿qu? modelo proporciona el mejor ajuste?
# (menores errores pasados (AIC,SIC, etc), mejores errores de predicción)
```

## 3.1 Estimación de un AR(0,0)

```
# Comienzo estimando sólo una constante

# librería tserie y forecast: estimo un modelo AR(0, 0)
modelo.arma <- tseries::arma(diff(log(pib.ts), k=4), order = c(0, 0))
```

```
## Warning in optim(coef, err, gr = NULL, hessian = TRUE, ...): one-dimensional optimization by Nelder-Mead is unreliable:
## use "Brent" or optimize() directly
```

```
summary(modelo.arma)
```

```
##
## Call:
## tseries::arma(x = diff(log(pib.ts), k = 4), order = c(0, 0))
##
## Model:
## ARMA(0,0)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.0223849 -0.0025704  0.0005926  0.0036352  0.0170913
##
## Coefficient(s):
##      Estimate Std. Error t value Pr(>|t|)
## intercept 0.0062622   0.0004577   13.68  <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Fit:
## sigma^2 estimated as 4.106e-05, Conditional Sum-of-Squares = 0.01, AIC = -1414.22
```

```
modelo.arma <- Arima(diff(log(pib.ts), k=4), order = c(0, 0, 0))
summary(modelo.arma)
```

```
## Series: diff(log(pib.ts), k = 4)
## ARIMA(0,0,0) with non-zero mean
##
## Coefficients:
##      mean
##      0.0063
## s.e. 0.0005
##
## sigma^2 estimated as 4.106e-05: log likelihood=708.61
## AIC=-1413.23 AICc=-1413.16 BIC=-1406.68
##
## Training set error measures:
##      ME      RMSE      MAE      MPE      MAPE
## Training set 5.549741e-18 0.006391095 0.004701215 281.7783 390.6845
##      MASE      ACF1
## Training set 1.03283 0.8995993
```

```
#resumen modelo AR(0)
summary(modelo.arma)
```

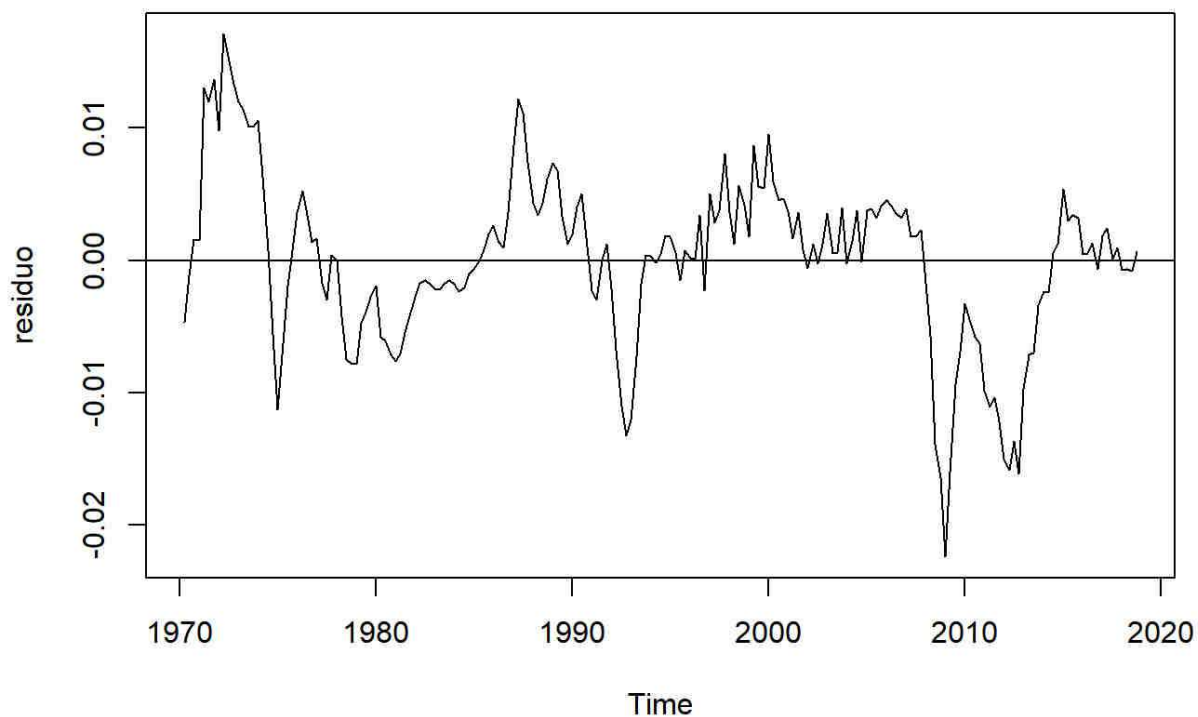
```
## Series: diff(log(pib.ts), k = 4)
## ARIMA(0,0,0) with non-zero mean
##
## Coefficients:
##      mean
##      0.0063
## s.e. 0.0005
##
## sigma^2 estimated as 4.106e-05: log likelihood=708.61
## AIC=-1413.23 AICc=-1413.16 BIC=-1406.68
##
## Training set error measures:
##      ME      RMSE      MAE      MPE      MAPE
## Training set 5.549741e-18 0.006391095 0.004701215 281.7783 390.6845
##      MASE      ACF1
## Training set 1.03283 0.8995993
```

```
# obtengo los residuos, si queda algo de dependencia
# en la serie sin modelizar, esa dependencia se reflejará
# en los residuos
```

```
residuo<-na.omit(residuals(modelo.arma))
length(residuo)
```

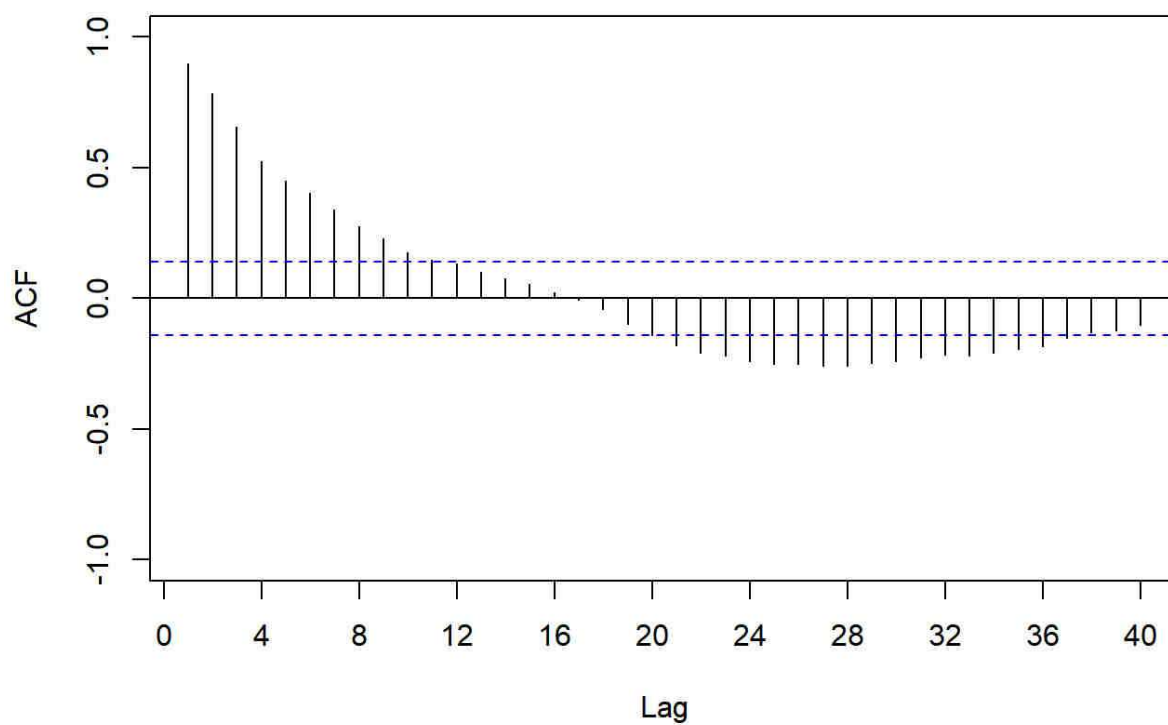
```
## [1] 195
```

```
plot(residuo, type="l")
abline(h=0)
```



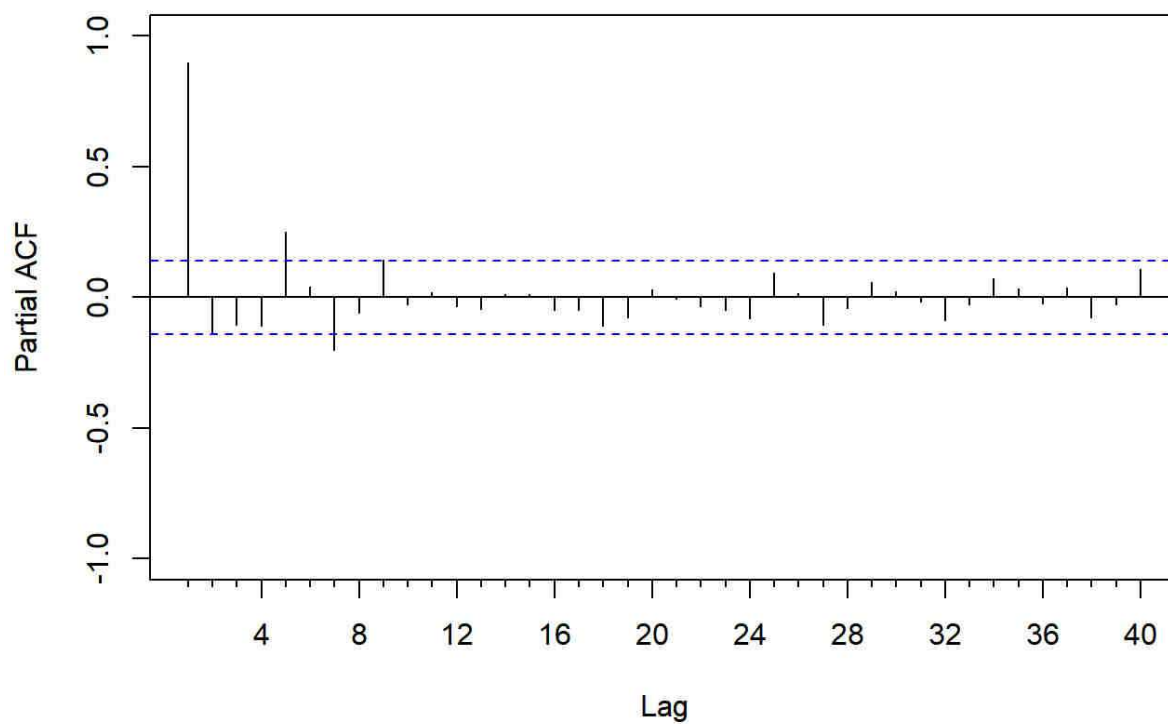
```
Acf(residuo, lag.max=40, ylim=c(-1,1))
```

### Series residuo



```
Pacf(residuo, lag.max=40, ylim=c(-1,1))
```

### Series residuo



## 3.2 Estimación de un AR(1,0)

```
# estimo ahora un modelo AR(1)
modelo.arma <- tseries::arma(diff(pib.ts, k=4), order = c(1, 0)) # librería forecast estimo un modelo AR(1)
summary(modelo.arma) #resumen modelo AR(1)
```

```
##
## Call:
## tseries::arma(x = diff(pib.ts, k = 4), order = c(1, 0))
##
## Model:
## ARMA(1,0)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.883345 -0.110586  0.002658  0.096358  0.543198
##
## Coefficient(s):
##           Estimate Std. Error t value Pr(>|t|)
## ar1           0.90127    0.03104   29.036  <2e-16 ***
## intercept     0.04239    0.01881    2.254   0.0242 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Fit:
## sigma^2 estimated as 0.04021, Conditional Sum-of-Squares = 7.76, AIC = -69.26
```

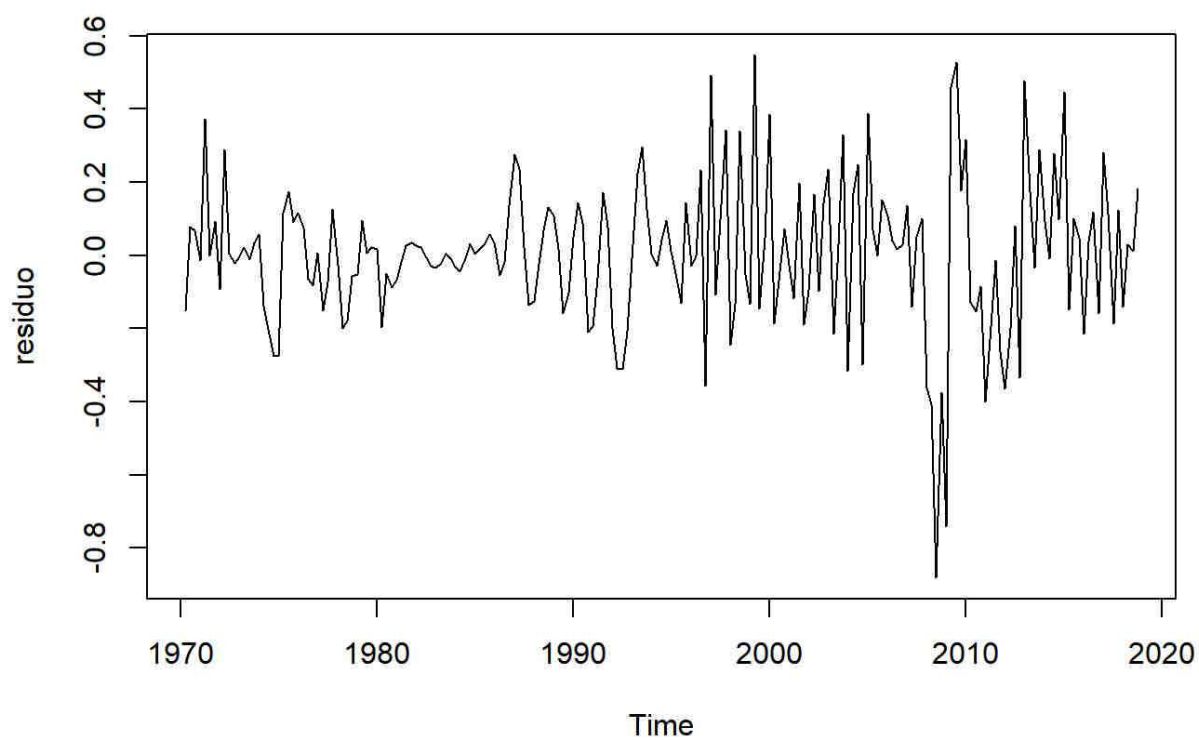
```
modelo.arma <- forecast::Arima(diff(pib.ts, k=4), order = c(1,0, 0)) # librería forecast estimo un modelo AR(1)
summary(modelo.arma) #resumen modelo AR(1)
```

```
## Series: diff(pib.ts, k = 4)
## ARIMA(1,0,0) with non-zero mean
##
## Coefficients:
##           ar1      mean
##           0.8993  0.3954
## s.e.      0.0304  0.1360
##
## sigma^2 estimated as 0.04034: log likelihood=36.5
## AIC=-66.99 AICc=-66.87 BIC=-57.17
##
## Training set error measures:
##           ME      RMSE      MAE      MPE      MAPE      MASE
## Training set 0.002567485 0.1998195 0.143534 22.53862 94.9375 0.4611776
##           ACF1
## Training set 0.1180273
```

```
#obtengo los residuos del modelo AR(1)
residuo<-na.omit(residuals(modelo.arma))
length(residuo)
```

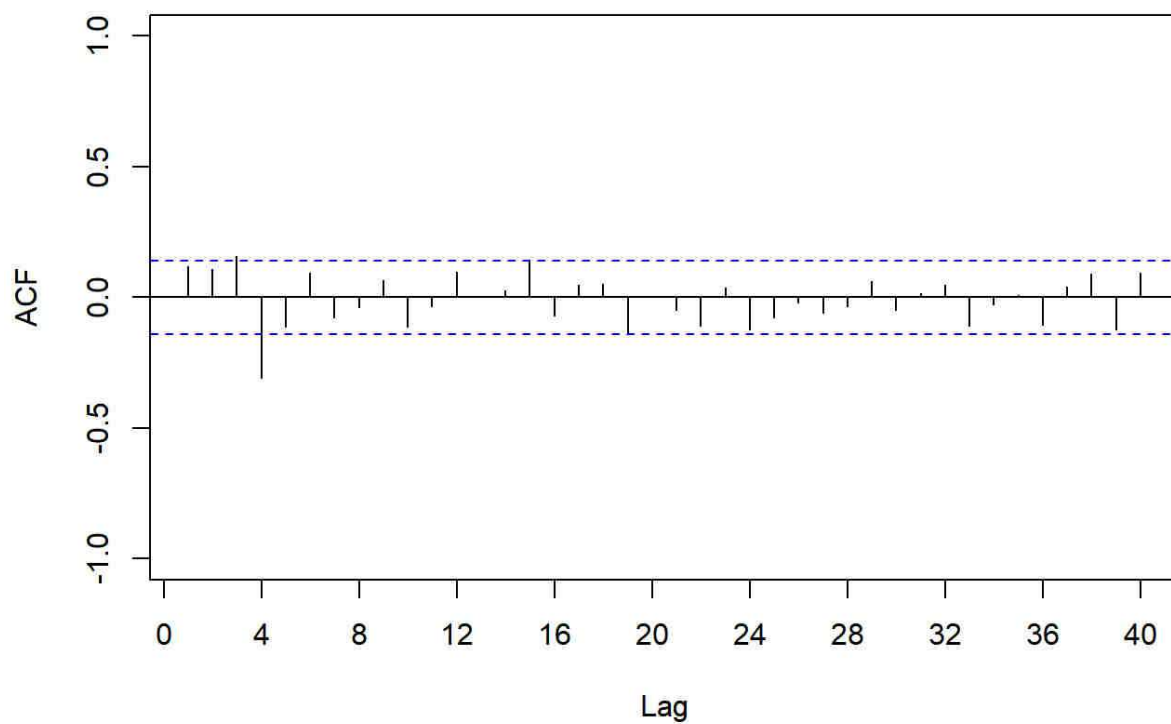
```
## [1] 195
```

```
plot(residuo, type="l")
```

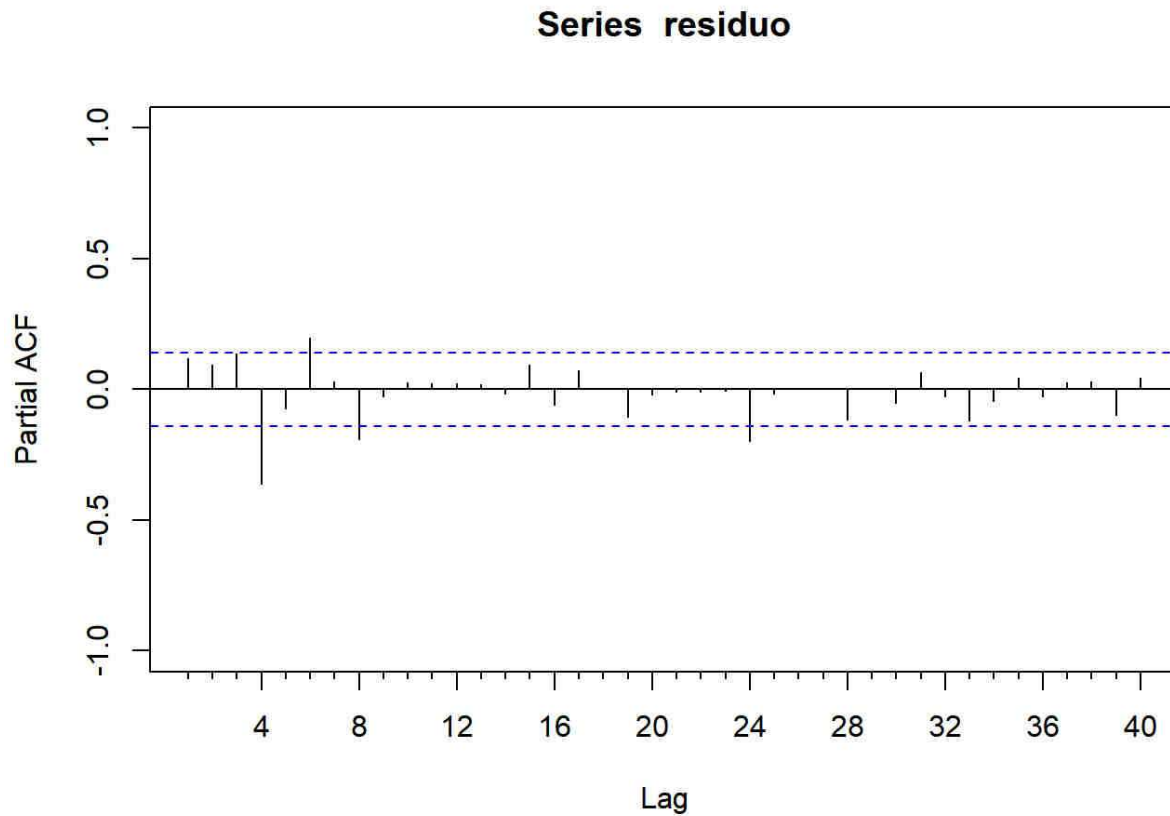


```
Acf(residuo, lag.max=40, ylim=c(-1,1))
```

### Series residuo



```
Pacf(residuo, lag.max=40, ylim=c(-1,1))
```



## 4 Estimación modelo automático: auto.arima

```
# auto.arima----  
##### Mejor ARIMA  
modelo.arma<-auto.arima(pib.ts,stepwise = FALSE, trace = TRUE) # librería forecast
```



```

##
## Fitting models using approximations to speed things up...
##
## ARIMA(0,1,0) : 361.5964
## ARIMA(0,1,0) with drift : 256.4338
## ARIMA(0,1,0)(0,0,1)[4] : 275.1247
## ARIMA(0,1,0)(0,0,1)[4] with drift : 211.0882
## ARIMA(0,1,0)(0,0,2)[4] : 253.6303
## ARIMA(0,1,0)(0,0,2)[4] with drift : 206.5749
## ARIMA(0,1,0)(1,0,0)[4] : 228.3613
## ARIMA(0,1,0)(1,0,0)[4] with drift : 203.4505
## ARIMA(0,1,0)(1,0,1)[4] : 226.8836
## ARIMA(0,1,0)(1,0,1)[4] with drift : 205.4796
## ARIMA(0,1,0)(1,0,2)[4] : 228.0784
## ARIMA(0,1,0)(1,0,2)[4] with drift : 207.5689
## ARIMA(0,1,0)(2,0,0)[4] : 226.4121
## ARIMA(0,1,0)(2,0,0)[4] with drift : 206.5989
## ARIMA(0,1,0)(2,0,1)[4] : 225.1104
## ARIMA(0,1,0)(2,0,1)[4] with drift : 207.4785
## ARIMA(0,1,0)(2,0,2)[4] : 227.2024
## ARIMA(0,1,0)(2,0,2)[4] with drift : 203.0938
## ARIMA(0,1,1) : 162.2056
## ARIMA(0,1,1) with drift : 79.15761
## ARIMA(0,1,1)(0,0,1)[4] : 98.27209
## ARIMA(0,1,1)(0,0,1)[4] with drift : 50.37163
## ARIMA(0,1,1)(0,0,2)[4] : 83.23214
## ARIMA(0,1,1)(0,0,2)[4] with drift : 47.92249
## ARIMA(0,1,1)(1,0,0)[4] : 72.11925
## ARIMA(0,1,1)(1,0,0)[4] with drift : 46.16609
## ARIMA(0,1,1)(1,0,1)[4] : 70.60792
## ARIMA(0,1,1)(1,0,1)[4] with drift : 48.26749
## ARIMA(0,1,1)(1,0,2)[4] : 72.67595
## ARIMA(0,1,1)(1,0,2)[4] with drift : 49.5316
## ARIMA(0,1,1)(2,0,0)[4] : 60.94272
## ARIMA(0,1,1)(2,0,0)[4] with drift : 45.28355
## ARIMA(0,1,1)(2,0,1)[4] : 60.47805
## ARIMA(0,1,1)(2,0,1)[4] with drift : 46.26635
## ARIMA(0,1,1)(2,0,2)[4] : 62.05699
## ARIMA(0,1,1)(2,0,2)[4] with drift : 45.1396
## ARIMA(0,1,2) : 87.98253
## ARIMA(0,1,2) with drift : 31.04637
## ARIMA(0,1,2)(0,0,1)[4] : 78.98553
## ARIMA(0,1,2)(0,0,1)[4] with drift : 30.12541
## ARIMA(0,1,2)(0,0,2)[4] : 64.94341
## ARIMA(0,1,2)(0,0,2)[4] with drift : 29.80408
## ARIMA(0,1,2)(1,0,0)[4] : 63.76556
## ARIMA(0,1,2)(1,0,0)[4] with drift : 28.82746
## ARIMA(0,1,2)(1,0,1)[4] : 53.12237
## ARIMA(0,1,2)(1,0,1)[4] with drift : 30.62823
## ARIMA(0,1,2)(1,0,2)[4] : 53.95115
## ARIMA(0,1,2)(1,0,2)[4] with drift : 29.95565
## ARIMA(0,1,2)(2,0,0)[4] : 54.7967
## ARIMA(0,1,2)(2,0,0)[4] with drift : 30.72958
## ARIMA(0,1,2)(2,0,1)[4] : 48.71587
## ARIMA(0,1,2)(2,0,1)[4] with drift : 31.37239
## ARIMA(0,1,3) : -13.66757
## ARIMA(0,1,3) with drift : -70.67947
## ARIMA(0,1,3)(0,0,1)[4] : -53.73154
## ARIMA(0,1,3)(0,0,1)[4] with drift : -88.20251
## ARIMA(0,1,3)(0,0,2)[4] : -63.10356
## ARIMA(0,1,3)(0,0,2)[4] with drift : -89.86658

```

##	ARIMA(0,1,3)(1,0,0)[4]	:	-58.58226
##	ARIMA(0,1,3)(1,0,0)[4] with drift	:	-87.6656
##	ARIMA(0,1,3)(1,0,1)[4]	:	-59.89953
##	ARIMA(0,1,3)(1,0,1)[4] with drift	:	-85.52146
##	ARIMA(0,1,3)(2,0,0)[4]	:	Inf
##	ARIMA(0,1,3)(2,0,0)[4] with drift	:	-91.95295
##	ARIMA(1,1,0)	:	-64.19896
##	ARIMA(1,1,0) with drift	:	-67.14732
##	ARIMA(1,1,0)(0,0,1)[4]	:	-91.77787
##	ARIMA(1,1,0)(0,0,1)[4] with drift	:	-92.46669
##	ARIMA(1,1,0)(0,0,2)[4]	:	-90.01191
##	ARIMA(1,1,0)(0,0,2)[4] with drift	:	-90.65611
##	ARIMA(1,1,0)(1,0,0)[4]	:	-86.12438
##	ARIMA(1,1,0)(1,0,0)[4] with drift	:	-86.9263
##	ARIMA(1,1,0)(1,0,1)[4]	:	-98.02768
##	ARIMA(1,1,0)(1,0,1)[4] with drift	:	-98.13175
##	ARIMA(1,1,0)(1,0,2)[4]	:	-96.53691
##	ARIMA(1,1,0)(1,0,2)[4] with drift	:	-96.57735
##	ARIMA(1,1,0)(2,0,0)[4]	:	-93.9984
##	ARIMA(1,1,0)(2,0,0)[4] with drift	:	-93.99737
##	ARIMA(1,1,0)(2,0,1)[4]	:	-93.33946
##	ARIMA(1,1,0)(2,0,1)[4] with drift	:	-93.13315
##	ARIMA(1,1,0)(2,0,2)[4]	:	-91.42986
##	ARIMA(1,1,0)(2,0,2)[4] with drift	:	-91.2417
##	ARIMA(1,1,1)	:	-64.01236
##	ARIMA(1,1,1) with drift	:	-67.77664
##	ARIMA(1,1,1)(0,0,1)[4]	:	-91.00986
##	ARIMA(1,1,1)(0,0,1)[4] with drift	:	-92.0253
##	ARIMA(1,1,1)(0,0,2)[4]	:	-89.11605
##	ARIMA(1,1,1)(0,0,2)[4] with drift	:	-90.08038
##	ARIMA(1,1,1)(1,0,0)[4]	:	-85.77754
##	ARIMA(1,1,1)(1,0,0)[4] with drift	:	-87.03724
##	ARIMA(1,1,1)(1,0,1)[4]	:	-97.53244
##	ARIMA(1,1,1)(1,0,1)[4] with drift	:	-98.02954
##	ARIMA(1,1,1)(1,0,2)[4]	:	-96.3711
##	ARIMA(1,1,1)(1,0,2)[4] with drift	:	-96.81737
##	ARIMA(1,1,1)(2,0,0)[4]	:	-93.77849
##	ARIMA(1,1,1)(2,0,0)[4] with drift	:	-94.12472
##	ARIMA(1,1,1)(2,0,1)[4]	:	-93.30379
##	ARIMA(1,1,1)(2,0,1)[4] with drift	:	-93.46652
##	ARIMA(1,1,2)	:	-69.49638
##	ARIMA(1,1,2) with drift	:	-75.05579
##	ARIMA(1,1,2)(0,0,1)[4]	:	-94.47609
##	ARIMA(1,1,2)(0,0,1)[4] with drift	:	-95.94589
##	ARIMA(1,1,2)(0,0,2)[4]	:	-92.44619
##	ARIMA(1,1,2)(0,0,2)[4] with drift	:	-93.88481
##	ARIMA(1,1,2)(1,0,0)[4]	:	-88.42854
##	ARIMA(1,1,2)(1,0,0)[4] with drift	:	-90.28567
##	ARIMA(1,1,2)(1,0,1)[4]	:	-100.8958
##	ARIMA(1,1,2)(1,0,1)[4] with drift	:	-101.8767
##	ARIMA(1,1,2)(2,0,0)[4]	:	-97.66853
##	ARIMA(1,1,2)(2,0,0)[4] with drift	:	-98.44294
##	ARIMA(1,1,3)	:	-89.04922
##	ARIMA(1,1,3) with drift	:	-98.66511
##	ARIMA(1,1,3)(0,0,1)[4]	:	-98.69539
##	ARIMA(1,1,3)(0,0,1)[4] with drift	:	-101.8612
##	ARIMA(1,1,3)(1,0,0)[4]	:	-96.75369
##	ARIMA(1,1,3)(1,0,0)[4] with drift	:	-101.8821
##	ARIMA(2,1,0)	:	-63.69417
##	ARIMA(2,1,0) with drift	:	-67.55545
##	ARIMA(2,1,0)(0,0,1)[4]	:	-90.62117
##	ARIMA(2,1,0)(0,0,1)[4] with drift	:	-91.70438

```

## ARIMA(2,1,0)(0,0,2)[4] : -88.6402
## ARIMA(2,1,0)(0,0,2)[4] with drift : -89.68086
## ARIMA(2,1,0)(1,0,0)[4] : -85.29945
## ARIMA(2,1,0)(1,0,0)[4] with drift : -86.76719
## ARIMA(2,1,0)(1,0,1)[4] : -97.84557
## ARIMA(2,1,0)(1,0,1)[4] with drift : -98.48085
## ARIMA(2,1,0)(1,0,2)[4] : -97.26943
## ARIMA(2,1,0)(1,0,2)[4] with drift : -97.81077
## ARIMA(2,1,0)(2,0,0)[4] : -93.39628
## ARIMA(2,1,0)(2,0,0)[4] with drift : -93.94779
## ARIMA(2,1,0)(2,0,1)[4] : -93.24911
## ARIMA(2,1,0)(2,0,1)[4] with drift : -93.57357
## ARIMA(2,1,1) : -61.47697
## ARIMA(2,1,1) with drift : -65.57325
## ARIMA(2,1,1)(0,0,1)[4] : -95.00929
## ARIMA(2,1,1)(0,0,1)[4] with drift : -99.46951
## ARIMA(2,1,1)(0,0,2)[4] : -92.90244
## ARIMA(2,1,1)(0,0,2)[4] with drift : -97.37809
## ARIMA(2,1,1)(1,0,0)[4] : -87.59248
## ARIMA(2,1,1)(1,0,0)[4] with drift : -91.60811
## ARIMA(2,1,1)(1,0,1)[4] : -100.8851
## ARIMA(2,1,1)(1,0,1)[4] with drift : Inf
## ARIMA(2,1,1)(2,0,0)[4] : -97.42571
## ARIMA(2,1,1)(2,0,0)[4] with drift : Inf
## ARIMA(2,1,2) : -59.37722
## ARIMA(2,1,2) with drift : -63.52854
## ARIMA(2,1,2)(0,0,1)[4] : -95.93236
## ARIMA(2,1,2)(0,0,1)[4] with drift : -98.91283
## ARIMA(2,1,2)(1,0,0)[4] : -90.16684
## ARIMA(2,1,2)(1,0,0)[4] with drift : -93.80017
## ARIMA(2,1,3) : -89.92015
## ARIMA(2,1,3) with drift : -98.27979
## ARIMA(3,1,0) : -62.85056
## ARIMA(3,1,0) with drift : -67.93913
## ARIMA(3,1,0)(0,0,1)[4] : -94.20729
## ARIMA(3,1,0)(0,0,1)[4] with drift : -96.16282
## ARIMA(3,1,0)(0,0,2)[4] : -92.08599
## ARIMA(3,1,0)(0,0,2)[4] with drift : -94.02978
## ARIMA(3,1,0)(1,0,0)[4] : -87.47314
## ARIMA(3,1,0)(1,0,0)[4] with drift : -90.07209
## ARIMA(3,1,0)(1,0,1)[4] : -99.70969
## ARIMA(3,1,0)(1,0,1)[4] with drift : -101.1776
## ARIMA(3,1,0)(2,0,0)[4] : -97.22023
## ARIMA(3,1,0)(2,0,0)[4] with drift : -98.82689
## ARIMA(3,1,1) : -66.623
## ARIMA(3,1,1) with drift : -67.3734
## ARIMA(3,1,1)(0,0,1)[4] : -91.2751
## ARIMA(3,1,1)(0,0,1)[4] with drift : -91.69046
## ARIMA(3,1,1)(1,0,0)[4] : -83.61661
## ARIMA(3,1,1)(1,0,0)[4] with drift : -91.47854
## ARIMA(3,1,2) : -65.28785
## ARIMA(3,1,2) with drift : -68.29172
##
## Now re-fitting the best model(s) without approximations...
##
##
##
## Best model: ARIMA(1,1,3)(1,0,0)[4] with drift

```

```
summary(modelo.arma)
```

```
## Series: pib.ts
## ARIMA(1,1,3)(1,0,0)[4] with drift
##
## Coefficients:
##          ar1      ma1      ma2      ma3      sar1      drift
##      0.8216  0.2123  0.3426  0.2958 -0.2464  0.4012
## s.e.  0.0783  0.1135  0.1098  0.1077  0.1119  0.1043
##
## sigma^2 estimated as 0.03314:  log likelihood=57.19
## AIC=-100.37  AICc=-99.77  BIC=-77.46
##
## Training set error measures:
##              ME      RMSE      MAE      MPE      MAPE
## Training set 0.001229919 0.1787532 0.1279339 0.001277229 0.1888453
##              MASE      ACF1
## Training set 0.0630616 -0.0195034
```

```
modelo.arma$aic
```

```
## [1] -100.3734
```

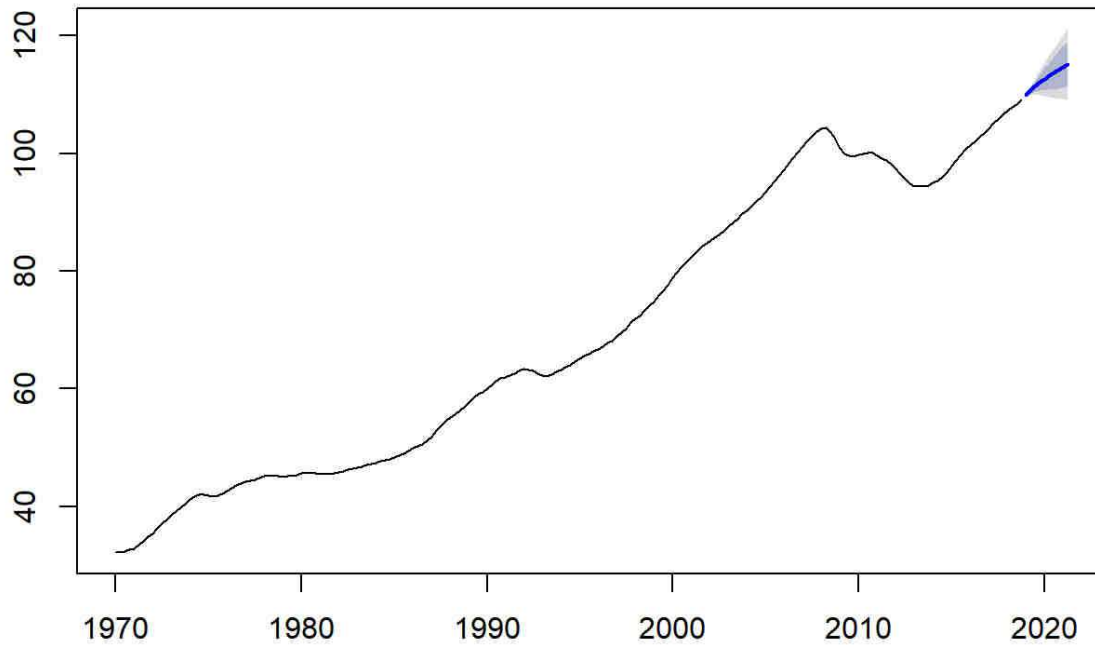
## 5 PREDICCIONES

```
forecast(modelo.arma, h=10)
```

```
##          Point Forecast    Lo 80    Hi 80    Lo 95    Hi 95
## 2019 Q1          109.9239 109.6906 110.1572 109.5671 110.2807
## 2019 Q2          110.6875 110.1587 111.2162 109.8789 111.4961
## 2019 Q3          111.4401 110.5204 112.3598 110.0335 112.8467
## 2019 Q4          112.0819 110.6860 113.4778 109.9471 114.2167
## 2020 Q1          112.6579 110.7928 114.5230 109.8055 115.5103
## 2020 Q2          113.1910 110.8719 115.5101 109.6442 116.7378
## 2020 Q3          113.6873 110.9390 116.4357 109.4841 117.8905
## 2020 Q4          114.1785 111.0296 117.3275 109.3626 118.9944
## 2021 Q1          114.6593 111.1262 118.1925 109.2558 120.0629
## 2021 Q2          115.1288 111.2255 119.0321 109.1593 121.0984
```

```
plot(forecast(modelo.arma, h=10))
```

Forecasts from ARIMA(1,1,3)(1,0,0)[4] with drift



## 6 Modelos Multivariantes: ARMAX

Cuando además del pasado de la serie, se incluye otra serie como regresor. Hay que utilizar la opción `xreg`

```
# Modelos Multivariantes: ARMAX ----

series<-ts.union(okun$T4pib.ts,okun$T4ocu.ts, dframe=TRUE)
series<-series[complete.cases(series),]

modelo.arma<-auto.arima(series[,1], xreg=series[,2],stepwise = FALSE, trace = TRUE) # librería forecast
```

```

##
## Fitting models using approximations to speed things up...
##
## ARIMA(0,1,0) : -1235.073
## Regression with ARIMA(0,1,0) errors : -1233.029
## ARIMA(0,1,1) : -1254.24
## Regression with ARIMA(0,1,1) errors : -1252.134
## ARIMA(0,1,2) : -1252.355
## Regression with ARIMA(0,1,2) errors : -1250.218
## ARIMA(0,1,3) : Inf
## Regression with ARIMA(0,1,3) errors : Inf
## ARIMA(0,1,4) : Inf
## Regression with ARIMA(0,1,4) errors : Inf
## ARIMA(0,1,5) : Inf
## Regression with ARIMA(0,1,5) errors : Inf
## ARIMA(1,1,0) : -1253.355
## Regression with ARIMA(1,1,0) errors : -1251.244
## ARIMA(1,1,1) : -1252.045
## Regression with ARIMA(1,1,1) errors : -1249.905
## ARIMA(1,1,2) : Inf
## Regression with ARIMA(1,1,2) errors : Inf
## ARIMA(1,1,3) : -1340.338
## Regression with ARIMA(1,1,3) errors : Inf
## ARIMA(1,1,4) : Inf
## Regression with ARIMA(1,1,4) errors : Inf
## ARIMA(2,1,0) : -1251.666
## Regression with ARIMA(2,1,0) errors : -1249.527
## ARIMA(2,1,1) : -1248.862
## Regression with ARIMA(2,1,1) errors : -1246.692
## ARIMA(2,1,2) : Inf
## Regression with ARIMA(2,1,2) errors : Inf
## ARIMA(2,1,3) : Inf
## Regression with ARIMA(2,1,3) errors : Inf
## ARIMA(3,1,0) : -1256.254
## Regression with ARIMA(3,1,0) errors : -1254.086
## ARIMA(3,1,1) : -1272.481
## Regression with ARIMA(3,1,1) errors : -1272.077
## ARIMA(3,1,2) : -1270.348
## Regression with ARIMA(3,1,2) errors : -1269.982
## ARIMA(4,1,0) : -1294.453
## Regression with ARIMA(4,1,0) errors : -1292.295
## ARIMA(4,1,1) : -1299.413
## Regression with ARIMA(4,1,1) errors : -1297.198
## ARIMA(5,1,0) : -1302.366
## Regression with ARIMA(5,1,0) errors : -1300.147
##
## Now re-fitting the best model(s) without approximations...
##
## ARIMA(0,1,0) : -1246.191
## Regression with ARIMA(0,1,0) errors : -1244.146
## ARIMA(0,1,1) : -1265.182
## Regression with ARIMA(0,1,1) errors : -1263.075
## ARIMA(0,1,2) : -1263.161
## Regression with ARIMA(0,1,2) errors : -1261.024
## ARIMA(0,1,3) : Inf
## Regression with ARIMA(0,1,3) errors : Inf
## ARIMA(0,1,4) : Inf
## Regression with ARIMA(0,1,4) errors : Inf
## ARIMA(0,1,5) : Inf
## Regression with ARIMA(0,1,5) errors : Inf
## ARIMA(1,1,0) : -1265.035

```

```

## Regression with ARIMA(1,1,0) errors : -1262.928
## ARIMA(1,1,1) : -1263.725
## Regression with ARIMA(1,1,1) errors : -1261.589
## ARIMA(1,1,2) : Inf
## Regression with ARIMA(1,1,2) errors : Inf
## ARIMA(1,1,3) : Inf
## Regression with ARIMA(1,1,3) errors : Inf
## ARIMA(1,1,4) : Inf
## Regression with ARIMA(1,1,4) errors : Inf
## ARIMA(2,1,0) : -1264.337
## Regression with ARIMA(2,1,0) errors : -1262.2
## ARIMA(2,1,1) : -1261.541
## Regression with ARIMA(2,1,1) errors : -1259.374
## ARIMA(2,1,2) : Inf
## Regression with ARIMA(2,1,2) errors : Inf
## ARIMA(2,1,3) : Inf
## Regression with ARIMA(2,1,3) errors : Inf
## ARIMA(3,1,0) : -1269.718
## Regression with ARIMA(3,1,0) errors : -1267.553
## ARIMA(3,1,1) : -1285.541
## Regression with ARIMA(3,1,1) errors : Inf
## ARIMA(3,1,2) : -1283.349
## Regression with ARIMA(3,1,2) errors : Inf
## ARIMA(4,1,0) : -1307.839
## Regression with ARIMA(4,1,0) errors : -1305.688
## ARIMA(4,1,1) : -1312.607
## Regression with ARIMA(4,1,1) errors : -1310.399
## ARIMA(5,1,0) : -1316.234
## Regression with ARIMA(5,1,0) errors : -1314.01
##
##
##
##
## Best model: Regression with ARIMA(5,1,0) errors

```

```
summary(modelo.arma)
```

```

## Series: series[, 1]
## Regression with ARIMA(5,1,0) errors
##
## Coefficients:
##      ar1      ar2      ar3      ar4      ar5      xreg
##      0.4055  0.0121 -0.0451 -0.5469  0.2724  0.5751
## s.e.  0.1005  0.0714  0.0738  0.0702  0.0833  0.0633
##
## sigma^2 estimated as 8.954e-06: log likelihood=665.51
## AIC=-1317.02  AICc=-1316.23  BIC=-1295.9
##
## Training set error measures:
##              ME              RMSE              MAE              MPE              MAPE
## Training set -2.196142e-05  0.002922683  0.002300765 -1.504268  13.7479
##              MASE              ACF1
## Training set 0.5562045 -0.01613948

```

```
modelo.arma$aic
```

```
## [1] -1317.018
```







Proyecto de innovación  
"Innova Docencia" de la UCM  
Convocatoria 2018-2019  
Proyecto número 275

**R-adaptación de la asignatura de métodos econométricos en economía y finanzas del  
Grado de Estadística Aplicada**

Facultad de Estudios Estadísticos  
Departamento de Economía Aplicada, Pública y Política

**Título**

***Practica 14 : Endogeneidad con un ejemplo sencillo: Estimación y  
Diagnosis***

**Correspondiente al**

**Capítulo 7 ENDOGENEIDAD Y EL ESTIMADOR DE VARIABLES INSTRUMENTALES**

- ¿En qué consiste el problema de la Endogeneidad?, ¿qué problemas origina?
  - Orígenes de la endogenidad: Omisión de alguna variable relevante (porque no es observable) que está correlacionada con alguna otra variable explicativa
  - Orígenes de la endogenidad: Errores en la medida
  - Orígenes de la endogenidad: Modelo de regresión con retardos y autocorrelación en los errores
  - Estimador de variables instrumentales y mínimos cuadrados en dos etapas (MC2E)
  - Contrastes de endogeneidad y restricciones de sobreestimación
- Practica 14 : Endogeneidad con un ejemplo sencillo: Estimación y Diagnosis***  
***Práctica 15 : Estimación del rendimiento de la educación para las mujeres casadas***

# Endogeneidad, el estimador de Variables instrumentales y MC2E: Ejemplode apuntes de clase

## Asignatura Métodos Econométricos en Economía y Finanzas

Profesores  
Loren Escot, Alicia Pérez-Alonso y Julio E. Sandubete

Dpto Economía Aplicada, Pública y Política  
Facultad Estudios Estadísticos - Universidad Complutense de Madrid

Curso académico 2018/2019

### Abstract

En esta práctica se revisará el problema de inconsistencia que genera la edogeniedad, esto es, cuando existe una covarianza no nula entre el término de error y alguno de los regresores del modelo. En esta primera práctica se muestra el ejemplo expuesto en los apuntes de clase.

- 1 Ejemplo Variables Instrumentales
- 2 Construcción de las Variables del Ejemplo
- 3 Estimación Inconsistente por MCO
- 4 Estimación por MC2E: Variable Instrumento Q
  - 4.1 con la librería AER
  - 4.2 Estimación de MC2E Manual
  - 4.3 Diagnósis para detectar Endogeneidad (y comprobar que se ha corregido con los instrumentos utilizados)
  - 4.4 Resumen de los Test DE DIAGNOSIS
    - 4.4.1 TEST DE ENDOGENEIDAD ( $H_0$ : EXOGENEIDAD)
    - 4.4.2 TEST DE DEBILIDAD DE LOS INSTRUMENTOS
    - 4.4.3 TEST DE ORTOGONALIDAD DE LOS INSTRUMENTOS
- 5 Estimación robusta de la Matriz de Var-Cov de los Estimadores MC2E
- 6 Sobreidentificación (dos o más instrumentos)

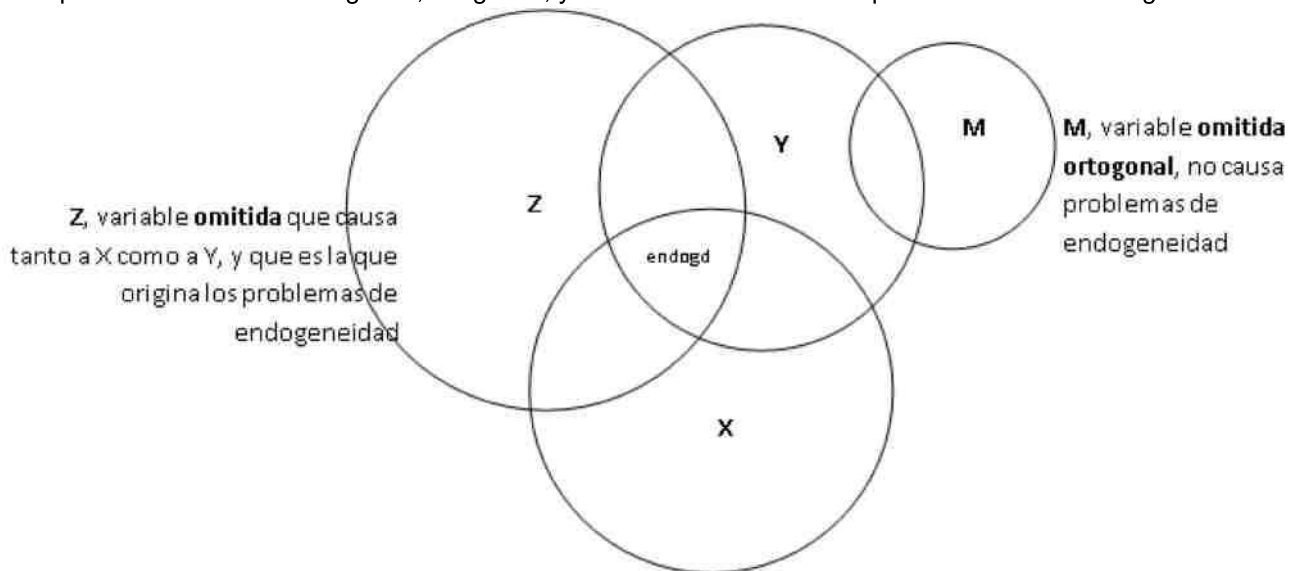
#=====

## 1 Ejemplo Variables Instrumentales

Vamos a seguir el ejemplo de los Apuntes en donde generábamos una serie de variables explicativas exógenas y otras endógenas para la estimación de una variable objetivo, y mostraremos cómo detectar el problema de la endogeneidad y las alternativas para solucionarlo.

Supongamos que la endogeneidad surge porque hay una causa común  $Z$  que no se observa. Al intentar hacer la regresión de  $X$  sobre  $Y$  aparece entonces el problema de la endogeneidad

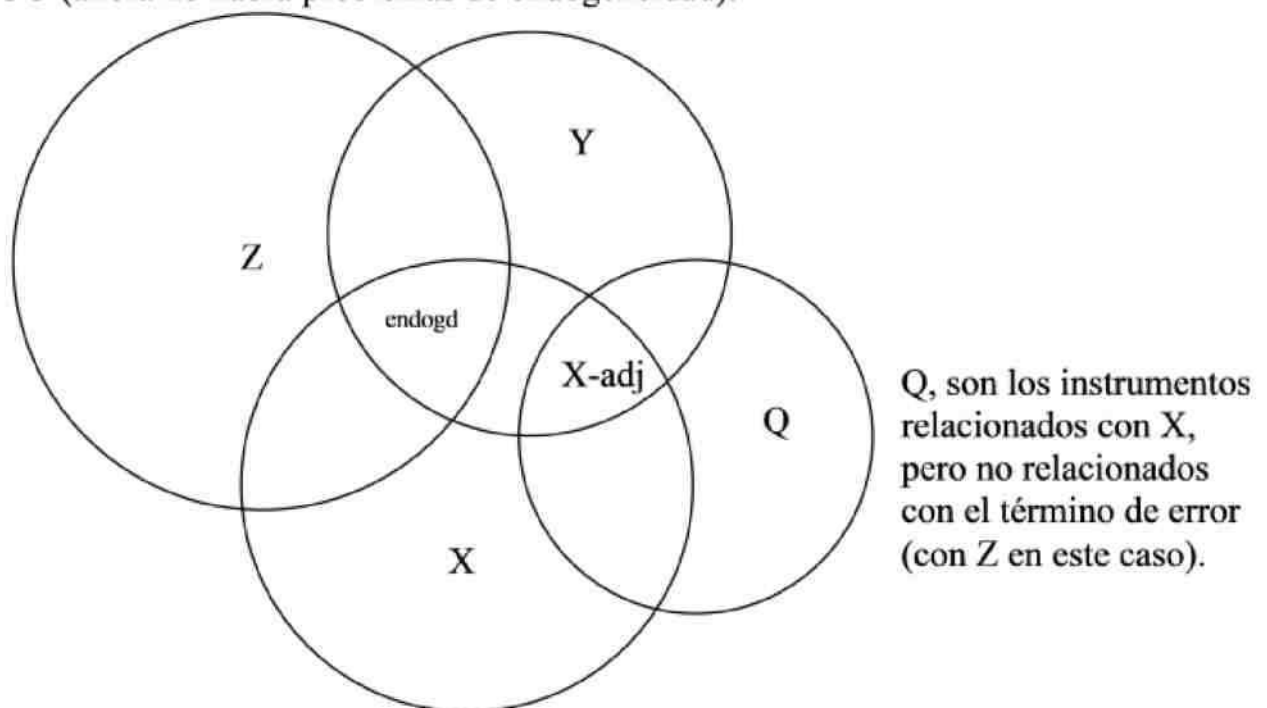
El esquema de Variables endógenas, exógenas, y los instrumentos válidos que se analizará es el siguiente



Para la Estimación consistente necesitamos un instrumento válido de X.

**Q instrumento:** relacionado con X, pero no relacionado con Z (M no sería un buen instrumento porque aunque está relacionado con Y y es ortogonal a Z, no está relacionado con X)

**MC2E(Mínimos Cuadrados en 2 Etapas):** Se ajusta X por MCO de Q, y se obtiene el valor ajustado de X por Q; luego se utiliza este valor ajustado para ajustar Y de nuevo por MCO (ahora no habrá problemas de endogeneidad).



## 2 Construcción de las Variables del Ejemplo

```

# Genero las Variables Independientes
set.seed(56)

# exógenas observables
x1 <- rnorm(1000,mean = 15,sd = 1)
x2 <- rnorm(1000,mean = 25,sd = 10)
x3 <- rnorm(1000,mean = 5,sd = 2)

# Z, variable omitida que causa tanto a X como a Y,
# y que es la que causa los problemas de endogeneidad

Z <- rnorm(1000,mean = 7,sd = 1)

# M, variable omitida ortogonal, no causa
# problemas de endogeneidad

M <- rnorm(1000,mean = 2,sd = 1)

# Q son los instrumentos, relacionados con X, pero no relacionados
# con el término de error (con Z en este caso).
Q <- rnorm(1000,mean = 7,sd = 10)

# En este ejemplo la variable Z va a ser la que causa la endogeneidad
# ya que es la causa común de X y de Y

# M es otra variable omitida pero que no causa endogeneidad

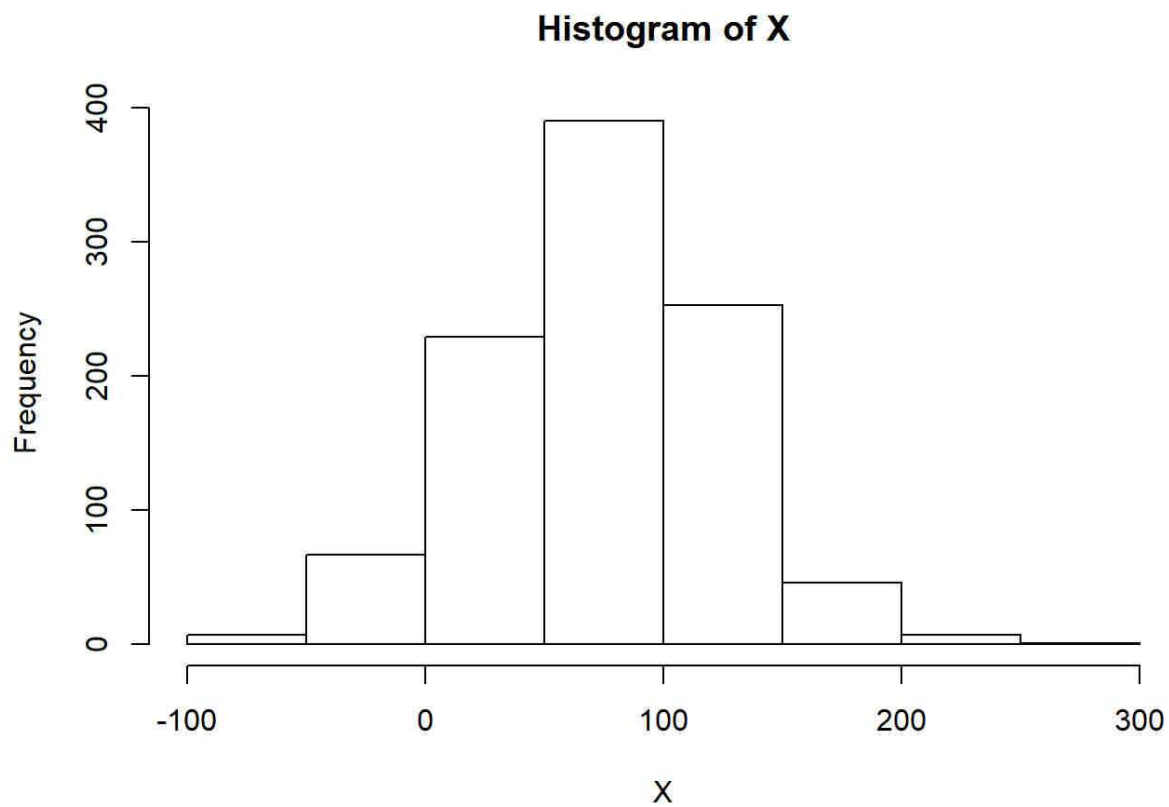
# La construcción de las variables ha sido la siguiente:

#  $x = 3 + 5*Q + 4*Z + 1.2*x_3 + \text{rnd1}$ 
#  $y = 2 + 0.4*x + 0.1*x_1 + 0.6*x_2 + 0.3*Z + 0.5*M + \text{rnd2}$ 

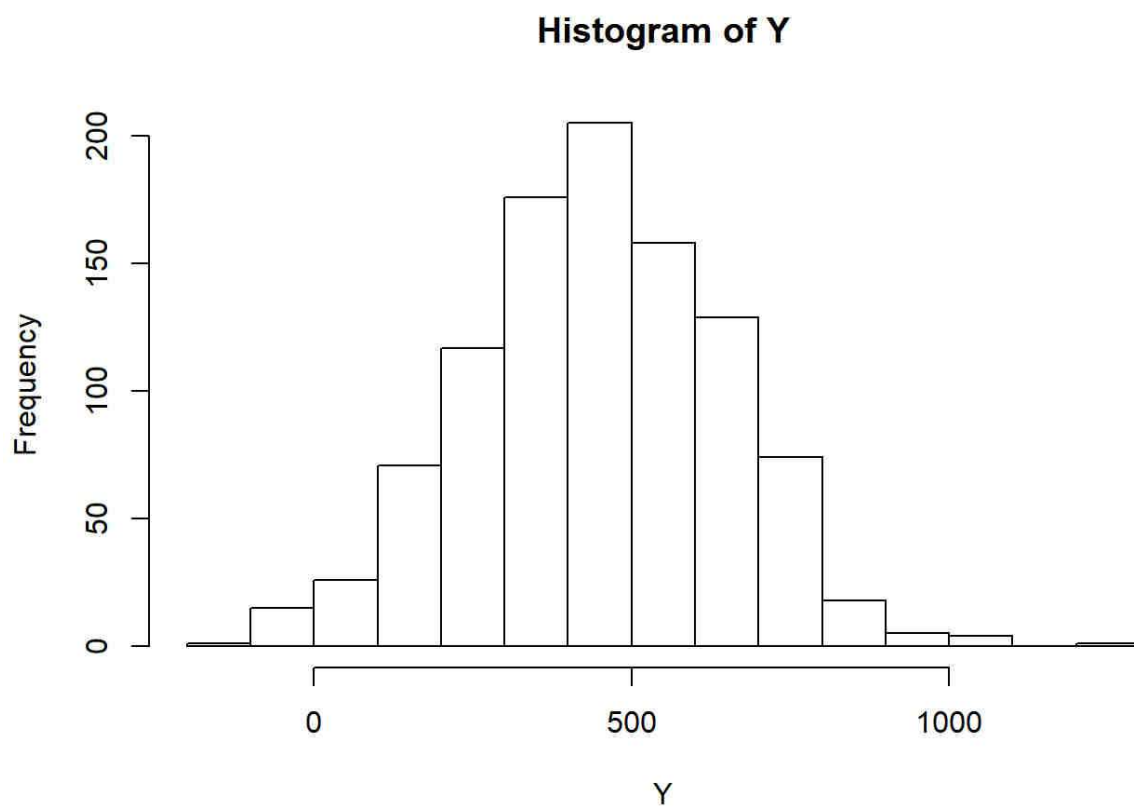
X<- 3 + 5*Q + 4*Z + 1.2*x3 + rnorm(1000,0,0.05)
Y<- 2 + 4*X + 2.1*x1 + 3*x2 + 5*Z + 3*M + rnorm(1000,0,0.07)

hist(X)

```



```
hist(Y)
```



## 3 Estimación Inconsistente por MCO

```
# Modelo inicial tenemos 3 posibles Vbles explicativas
modelo.lm<-lm(Y~ x1 + x2 + X)
summary(modelo.lm)
```

```
##
## Call:
## lm(formula = Y ~ x1 + x2 + X)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -22.2195  -4.1605  -0.0821   3.7960  19.4169
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 42.526639   2.838914   14.98  <2e-16 ***
## x1           2.036033   0.187069   10.88  <2e-16 ***
## x2           3.032685   0.019209  157.88  <2e-16 ***
## X            4.008996   0.003747 1070.01  <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 5.842 on 996 degrees of freedom
## Multiple R-squared:  0.9991, Adjusted R-squared:  0.9991
## F-statistic: 3.895e+05 on 3 and 996 DF,  p-value: < 2.2e-16
```

## 4 Estimación por MC2E: Variable Instrumento Q

```
# Modelo MC2E : Variable Instrumental Q
```

### 4.1 con la librería AER

```
suppressMessages(library(AER))

modelo.2lm<-ivreg(Y~ x1 + x2 + X | x1 + x2 + Q)
summary(modelo.2lm)
```

```
##
## Call:
## ivreg(formula = Y ~ x1 + x2 + X | x1 + x2 + Q)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -22.3803  -4.0765  -0.2041   3.7150  19.7396
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 43.100488   2.846497   15.14  <2e-16 ***
## x1           2.040599   0.187562   10.88  <2e-16 ***
## x2           3.032392   0.019260  157.45  <2e-16 ***
## X            4.000408   0.003774 1059.92  <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 5.858 on 996 degrees of freedom
## Multiple R-Squared:  0.9991, Adjusted R-squared:  0.9991
## Wald test: 3.823e+05 on 3 and 996 DF, p-value: < 2.2e-16
```

## 4.2 Estimación de MC2E Manual

```
# estimación manual
modelo.2lm_1<-lm(X~x1 + x2 + Q)
summary(modelo.2lm_1)
```

```
##
## Call:
## lm(formula = X ~ x1 + x2 + Q)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -13.7763  -3.3894   0.0473   3.3911  13.6605
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 35.02644   2.31474   15.132  <2e-16 ***
## x1           0.06920   0.15299   0.452   0.651
## x2           0.03631   0.01571   2.311   0.021 *
## Q            5.00001   0.01539 324.828  <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 4.778 on 996 degrees of freedom
## Multiple R-squared:  0.9907, Adjusted R-squared:  0.9906
## F-statistic: 3.518e+04 on 3 and 996 DF, p-value: < 2.2e-16
```

```
Xfit<-fitted(modelo.2lm_1)

modelo.2lm_2<-lm(Y ~x1 + x2 + Xfit)
summary(modelo.2lm_2)
```

```
##
## Call:
## lm(formula = Y ~ x1 + x2 + Xfit)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -63.679 -16.551   0.163  16.927  68.292
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  43.10049   11.57512   3.724 0.000207 ***
## x1           2.04060    0.76271   2.675 0.007585 **
## x2           3.03239    0.07832  38.719 < 2e-16 ***
## Xfit         4.00041    0.01535 260.650 < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 23.82 on 996 degrees of freedom
## Multiple R-squared:  0.9858, Adjusted R-squared:  0.9858
## F-statistic: 2.312e+04 on 3 and 996 DF,  p-value: < 2.2e-16
```

## 4.3 Diagnósis para detectar Endogeneidad (y comprobar que se ha corregido con los instrumentos utilizados)

### Resumen de los Test

- El instrumento NO debe ser débil: **WEAK test** (H0: el instrumento es débil (no correlacionado con X, y por tanto no válido))
  - También se podría haber estudiado el coeficiente de correlación entre variable endógena X y el instrumento
- El Test de Endogeneidad **Durbin-Wu-Hausman** (H0: Exogeneidad) Para saber si había un problema de Endogeneidad que hemos solucionado con nuestros instrumentos debemos rechazarla Hipótesis nula
- TEST de **Sargant** DE ORTOGONALIDAD de los Instrumentos (Test of overidentifying restrictions) H0: Todos los instrumentos son ortogonales (Este test de Sargan sólo puede aplicarse cuando existe sobreidentificación, es decir más de un instrumento)

```
#

# para incluir los diagnósticos
modelo.2lm<-ivreg(Y~ x1 + x2 + X | x1 + x2 + Q)
summary(modelo.2lm, diagnostics=TRUE)
```



```
##
## Call:
## ivreg(formula = Y ~ x1 + x2 + X | x1 + x2 + Q)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -22.3803  -4.0765  -0.2041   3.7150  19.7396
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 43.100488   2.846497   15.14  <2e-16 ***
## x1           2.040599   0.187562   10.88  <2e-16 ***
## x2           3.032392   0.019260  157.45  <2e-16 ***
## X            4.000408   0.003774 1059.92  <2e-16 ***
##
## Diagnostic tests:
##              df1 df2 statistic p-value
## Weak instruments    1 996   105514  <2e-16 ***
## Wu-Hausman          1 995    1260  <2e-16 ***
## Sargan              0 NA         NA      NA
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 5.858 on 996 degrees of freedom
## Multiple R-Squared: 0.9991, Adjusted R-squared: 0.9991
## Wald test: 3.823e+05 on 3 and 996 DF, p-value: < 2.2e-16
```

## 4.4 Resumen de los Test DE DIAGNOSIS

### 4.4.1 TEST DE ENDOGENEIDAD (H0: EXOGENEIDAD)

```
# 1: TEST DE ENDOGENEIDAD (H0: EXOGENEIDAD)

# Test de Endogeneidad 1: Test de Hausman

# (Ho: exogeneidad, Ha: endogeneidad)
# * estimador1:consistente bajo H0 y Ha
# * estimador2:eficiente bajo H0 pero INCONSISTENTE bajo Ha
# EL test contrasta la Ho: de que el estimador2 es eficiente, es decir, la hip?tesis de exogeneidad.
# H0: No existen diferencias entre el estimador1 y el estimador2
# Si los dos estimadores son muy diferentes entonces se rechaza la hip?tesis nula y tenemos un problema de endogeneidad
```

```
cf_diff <- coef(modelo.2lm) - coef(modelo.lm)
vc_diff <- vcov(modelo.2lm) - vcov(modelo.lm)
x2_diff <- as.vector(t(cf_diff) %*% solve(vc_diff) %*% cf_diff)
pchisq(x2_diff, df = 1, lower.tail = FALSE) #grados de libertad= posibles vbles endógenas
```

```
## [1] 2.241506e-79
```

```
#Test de Endogeneidad 2: Test de Durbin-Wu-Hausman
# Es una variante del Test de Exogeneidad de Hausman
# (Ho: exogeneidad, Ha: endogeneidad)

#Este contraste de endogeneidad de D-Wu-Hausman (Ho Exogeneidad) consiste
# básicamente en contrastar si la parte no explicada por el instrumento
# (residuo de la primera regresión) es significativa en la segunda
```

## 4.4.2 TEST DE DEBILIDAD DE LOS INSTRUMENTOS

```
# 2 TEST DE DEBILIDAD DE LOS INSTRUMENTOS
# Los instrumentos tienen que estar correlacionados con la variable endógena
# En ocasiones se acepta la hipótesis de exogeneidad,
# pero porque los instrumentos no son válidos, son débiles,
# en el sentido de que están muy poco relacionados con la variable X
# En la primera etapa de MC2E se muestra con el test t si existe o no autocorrelación
# además se puede analizar directamente la correlación, u otras pruebas
# que contrastan si existe relación o no (si explican globalmente (test F de Wald))

#Para que un instrumento sea válido tiene que estar correlacionado con X pero no con el término de error
#Los tests de debilidad del instrumento miran la correlación con la variable dependiente

# Test1: correlaciones entre los regresores endógenos (X) y los instrumentos(Z)
#H0: el instrumento es débil (no correlacionado con X, y por tanto no válido)
cor.test(X,Q)
```

```
##
## Pearson's product-moment correlation
##
## data: X and Q
## t = 324.25, df = 998, p-value < 2.2e-16
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
## 0.9946662 0.9958365
## sample estimates:
## cor
## 0.9952875
```

```
# Test2 de debilidad del instrumento
#H0: el instrumento es débil (no correlacionado con X, y por tanto no válido)

# WeAK test Es un test de F de Wald de significación conjunta de la primera etapa
# (tienen que ser significativos o distintos de cero, para ello
# el estadístico F tiene que ser grande, mayor que el valor tabulado bajo la hipótesis nula (90% con
fianza).
# Entonces si se acepta la H0 de el instrumento no es válido por incorrelacionado con X
```

## 4.4.3 TEST DE ORTOGONALIDAD DE LOS INSTRUMENTOS

```
# Test3: TEST DE ORTOGONALIDAD (Test of overidentifying restrictions).

# Los test de overidentification o de ortogonalidad
# constatan que no exista correlaci?n entre instrumentos y t?rmino de error

#Test de ortogonalidad de los instrumentos frente al t?rmino de error

# Ho: Todos los instrumentos son ortogonales
#(test de Sargan cuando existe sobreidentificaci?n
#y todos los instrumentos proporcionan resultados similares),

# si se rechaza la hipotesis nula de ortogonalidad al menos un instrumento no es v?lido.
# Por el contrario si se acepta la H0: todos los instrumentos son ortogonales,
# es decir son v?lidos
```

## 5 Estimaci3n robusta de la Matriz de Var-Cov de los Estimadores MC2E

```
# si sospechamos que hay heterocedasticidad
library(sandwich)
summary(modelo.2lm, vcov=sandwich,diagnostics=TRUE)
```

```
##
## Call:
## ivreg(formula = Y ~ x1 + x2 + X | x1 + x2 + Q)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -22.3803  -4.0765  -0.2041   3.7150  19.7396
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 43.100488   3.001568   14.36  <2e-16 ***
## x1           2.040599   0.196253   10.40  <2e-16 ***
## x2           3.032392   0.019189  158.03  <2e-16 ***
## X            4.000408   0.003738 1070.07  <2e-16 ***
##
## Diagnostic tests:
##              df1 df2 statistic p-value
## Weak instruments    1 996    108093  <2e-16 ***
## Wu-Hausman          1 995     1202  <2e-16 ***
## Sargan              0 NA         NA     NA
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 5.858 on 996 degrees of freedom
## Multiple R-Squared: 0.9991, Adjusted R-squared: 0.9991
## Wald test: 3.881e+05 on 3 and 996 DF, p-value: < 2.2e-16
```

## 6 Sobreidentificaci3n (dos o m3s instrumentos)

```
# Modelo MC2E: Variables Instrumentales Q, x3
modelo.2lm<-ivreg(Y~ x1 + x2 + X | x1 + x2 + Q+ x3)
summary(modelo.2lm, vcov=sandwich,diagnostics=TRUE)
```

```
##
## Call:
## ivreg(formula = Y ~ x1 + x2 + X | x1 + x2 + Q + x3)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -22.3758  -4.0910  -0.1924   3.7080  19.7305
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 43.084388   3.001885   14.35  <2e-16 ***
## x1           2.040471   0.196253   10.40  <2e-16 ***
## x2           3.032401   0.019184  158.06  <2e-16 ***
## X            4.000649   0.003726 1073.83  <2e-16 ***
##
## Diagnostic tests:
##              df1 df2 statistic p-value
## Weak instruments    2 995 74717.944  <2e-16 ***
## Wu-Hausman          1 995 2933.944  <2e-16 ***
## Sargan              1  NA     1.539   0.215
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 5.857 on 996 degrees of freedom
## Multiple R-Squared: 0.9991, Adjusted R-squared: 0.9991
## Wald test: 3.912e+05 on 3 and 996 DF, p-value: < 2.2e-16
```



Proyecto de innovación  
“Innova Docencia” de la UCM  
Convocatoria 2018-2019  
Proyecto número 275

**R-adaptación de la asignatura de métodos econométricos en economía y finanzas del  
Grado de Estadística Aplicada**

Facultad de Estudios Estadísticos  
Departamento de Economía Aplicada, Pública y Política

**Título**

***Práctica 15 : Estimación del rendimiento de la educación para las  
mujeres casadas***

**Correspondiente al**

**Capítulo 7 ENDOGENEIDAD Y EL ESTIMADOR DE VARIABLES INSTRUMENTALES**

- ¿En qué consiste el problema de la Endogeneidad?, ¿qué problemas origina?
  - Orígenes de la endogeneidad: Omisión de alguna variable relevante (porque no es observable) que está correlacionada con alguna otra variable explicativa
  - Orígenes de la endogeneidad: Errores en la medida
  - Orígenes de la endogeneidad: Modelo de regresión con retardos y autocorrelación en los errores
  - Estimador de variables instrumentales y mínimos cuadrados en dos etapas (MC2E)
  - Contrastes de endogeneidad y restricciones de sobreestimación
- Práctica 14 : Endogeneidad con un ejemplo sencillo: Estimación y Diagnóstico***  
***Práctica 15 : Estimación del rendimiento de la educación para las mujeres casadas***

# Ejemplo: Estimación del rendimiento de la educación para las mujeres casadas

© 2019 *Métodos Econométricos en Economía y Finanzas*

Escot Mangas, Lorenzo & Pérez Alonso, Alicia & Sandubete Galán, Julio E.

Facultad de Estudios Estadísticos, Universidad Complutense de Madrid

30 junio 2019

## El problema de las variables omitidas

Consideramos el siguiente modelo para datos de sección cruzada

$$y_i = \beta x_i + \eta_i + v_i, \quad i = 1, \dots, N$$

- $y_i$  : (log) salario
- $x_i$  : educación
- $\beta$  : parámetro estructural
- $\eta_i$  : habilidad y/o motivación del individuo
- $v_i$  : error o perturbación estocástica fuera del control del individuo

En el enfoque moderno  $\eta_i$  se denomina efecto inobservable, componente inobservable, variable latente, heterogeneidad inobservable o heterogeneidad individual.

La identificación del parámetro  $\beta$  requiere

$$\mathbb{E}[v_i | x_i, \eta_i] = 0 \quad (\text{C1})$$

- Si  $\eta_i$  observable, identificamos  $\beta$  en una regresión de  $y$  sobre  $x$  y  $\eta$ .
- Si  $\eta_i$  inobservable, identificamos  $\beta$  en una regresión de  $y$  sobre  $x$  si

$$\mathbb{E}[\eta_i + v_i | x_i] = 0 \quad (\text{C2})$$

Así que además de (C1), vamos a necesitar

$$\mathbb{E}[\eta_i | x_i] = 0$$

**Caso 1:** Regresión de  $y_i$  sobre  $x_i$  y  $\eta_i$ , dado (C1):

Predictor Lineal Óptimo:

```
$$ \text{PLO: } \mathbb{E}^*[y_i | x_i, \eta_i] = \mathbb{E}[y_i | x_i, \eta_i] = \beta x_i + \eta_i $$
```

Condiciones de ortogonalidad:

$$\mathbb{E}[v_i] = 0$$

$$\mathbb{E}[x_i v_i] = \text{Cov}(x_i, v_i) = 0$$

$$\mathbb{E}[x_i \eta_i] = \text{Cov}(x_i, \eta_i) = 0$$

De la segunda condición de ortogonalidad obtenemos

$$\text{Cov}(x_i, v_i) = \text{Cov}(x_i, y_i - \beta x_i - \eta_i) = 0$$

$$\beta = \frac{\text{Cov}(x_i, y_i) - \text{Cov}(x_i, \eta_i)}{\text{Var}(x_i)}$$

**Caso 2:** Regresión de  $y_i$  sobre  $x_i$  donde se omite  $\eta_i$  (omisión de variable relevante). En este caso, el modelo de regresión alternativo sería:

$$y_i = \gamma x_i + \epsilon_i, \quad i = 1, \dots, N$$

donde  $\epsilon_i = \eta_i + v_i$ . Identificamos  $\gamma$  en una regresión de  $y$  sobre  $x$ .

Predictor Lineal Óptimo:

$$\text{PLO: } E[y_i | x_i, \eta_i] = \gamma x_i$$

Condiciones de ortogonalidad:

$$E[y_i - \gamma x_i] = 0$$

$$E[x_i(y_i - \gamma x_i)] = Cov(x_i, y_i - \gamma x_i) = 0$$

De la segunda condición de ortogonalidad obtenemos

$$\gamma = \frac{Cov(x_i, y_i)}{Var(x_i)}$$

Por lo tanto,  $\gamma = \beta$  si y solo si  $Cov(x_i, \eta_i) = 0$ .

Sesgo de especificación por omisión de variable relevante:

- Si  $Cov(x_i, \eta_i) > 0 \implies$  sobre-estimamos el efecto de la educación  $\gamma$
- Si  $Cov(x_i, \eta_i) < 0 \implies$  subestimamos el efecto de la educación  $\gamma$

Por tanto, la heterogeneidad inobservables es un problema solo si está correlacionada con el resto de regresores observables, en ese caso se dice que los regresores son endógenos. Si  $E[\eta_i | x_i] \neq 0$ , en general  $E[x_i \eta_i] \neq 0$  y el estimador de mínimos cuadrados ordinarios (OLS por sus siglas en inglés) es inconsistente; será sesgado y ese sesgo no tiende a aislarse asintóticamente con el tamaño muestral.

Para obtener estimadores consistentes (insesgados asintóticamente) habría que utilizar otros métodos de estimación como:

- Variables instrumentales (IV por sus siglas en inglés)
- Mínimos cuadrados en dos etapas (2SLS por sus siglas en inglés)
- Método generalizado de los momentos (GMM por sus siglas en inglés)

## Estimación por el método de variables instrumentales

Si existiese un instrumento  $z_i$  para el cual se cumple

$$E[\eta_i + v_i | z_i] = 0 \tag{C3}$$

podríamos identificar  $\beta$  y estimarla usando el estimador de variables instrumentales.

$$\hat{\beta}_{IV} = \frac{Cov(z_i, y_i)}{Cov(z_i, x_i)}$$

Para que un instrumento  $z$  sea válido tiene que cumplir dos condiciones:

- **Relevancia** del instrumento: Tiene que estar correlacionado con  $x$  para que el denominador no sea cero,  $E[z_i x_i] \neq 0$ .
- **Exogeneidad** del instrumento: Tiene que estar incorrelacionado con el término de error,  $E[z_i(\eta_i + v_i)] \neq 0$  (se deriva de (C3)).

Consideramos el caso general de un modelo de regresión lineal múltiple con regresores endógenos:

$$Y_i = X_i' \beta + \epsilon_i, \quad i = 1, \dots, N, \quad \mathbb{E}[X_i \epsilon_i] \neq 0$$

El método de estimación por variables instrumentales consiste en encontrar un conjunto de  $q \geq k$  variables instrumentales  $Z_i$  que sean exógenas y relevantes. En el caso más sencillo  $q = k$  el modelo está exactamente identificado, y si  $q > k$  el modelo está sobreidentificado.

El estimador de IV se obtiene como un estimador del método de los momentos basado en la condición de  $\mathbb{E}[Z_i \epsilon_i] = 0$ .

$$\mathbb{E}[Z_i \epsilon_i] = \mathbb{E}[Z_i (Y_i - X_i' \beta)] = 0$$

$$\mathbb{E}[Z_i Y_i] = \mathbb{E}[Z_i X_i'] \beta$$

Utilizando el principio de analogía,

$$\hat{\beta}_{IV} = (Z' X)^{-1} (Z' Y)$$

Si los instrumentos son válidos, este estimador será consistente.

Comentarios:

1. En la mayoría de las aplicaciones no todos los regresores son endógenos. En ese caso, los regresores que sean exógenos se incluyen en el conjunto de instrumentos. Veamos un ejemplo:

$$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \epsilon_i$$

donde  $\mathbb{E}[x_{1i} \epsilon_i] = 0$ ,  $\mathbb{E}[x_{2i} \epsilon_i] \neq 0$ , y sea  $z_i$  un instrumento exógeno  $\mathbb{E}[z_i \epsilon_i] = 0$  y relevante  $\mathbb{E}[z_i x_{2i}] \neq 0$ . El vector de instrumentos será en este caso:  $Z_i = [1 \quad x_{1i} \quad z_i]'$ .

2. El estimador OLS es un caso particular del estimador IV cuando los instrumentos utilizados son las variables explicativas.

## Estimación por el método de mínimos cuadrados en dos etapas

El nombre de 2SLS se debe a que el estimador se obtiene mediante una estimación por mínimos cuadrados ordinarios en dos etapas.

1.<sup>a</sup> etapa: Regresamos por separado cada uno de los regresores potencialmente endógenos sobre todos los instrumentos y todos los regresores exógenos de la ecuación original para obtener su valor ajustado.

2.<sup>a</sup> etapa: Estimamos por mínimos cuadrados ordinarios la ecuación original sustituyendo los regresores endógenos por sus valores ajustados estimados en la primera etapa.

Uno de los inconvenientes de este cálculo manual es que el estimador de la matriz de varianzas-covarianzas del vector de parámetros  $\beta$  no es válido.

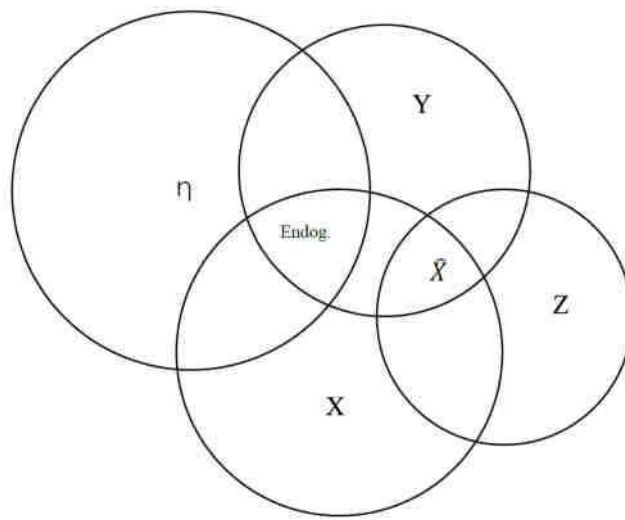
Si usamos el modelo del ejemplo anterior:

1.<sup>a</sup> etapa: Hacemos una regresión de  $x_{2i}$  sobre una constante,  $x_{1i}$  y  $z_i$  y obtenemos  $\hat{x}_{2i}$ .

2.<sup>a</sup> etapa: Regresamos  $y_i$  sobre una constante,  $x_{1i}$  y  $\hat{x}_{2i}$ .

También podemos verlo gráficamente. La endogeneidad surge porque hay una causa común  $\eta$  que no se observa pero que causa tanto a  $X$  como a  $Y$ . Sea  $Z$  el conjunto de instrumentos relacionados con  $X$  pero no con el término de error, la variable omitida  $\eta$  en este caso. Mediante el método 2SLS, regresamos  $X$  sobre  $Z$  para obtener su valor ajustado  $\hat{X}$  y después utilizamos  $\hat{X}$  para ajustar  $Y$  por mínimos cuadrados ordinarios, ya que ahora no habrá problemas de endogeneidad.





## El contraste de Wu-Hausman

Es un contraste de exogeneidad o de ausencia de correlación entre las variables explicativas y el término de error del modelo de regresión. Más concretamente:  $H_0: E[\epsilon_i | X_i] = 0$  vs  $H_1: E[\epsilon_i | X_i] \neq 0$ . El contraste de Wu-Hausman requiere la estimación del vector de parámetros  $\beta$  y la de su matriz de varianzas-covarianzas  $V$ . El estimador OLS es consistente y asintóticamente eficiente bajo  $H_0$  pero será inconsistente bajo  $H_1$ . El estimador de IV será consistente siempre, bajo  $H_0$  y  $H_1$ . Por tanto, si existe endogeneidad solamente IV es consistente.

El estadístico de contraste es el siguiente:

$$h = (\hat{\beta}_{OLS} - \hat{\beta}_{IV})'(\hat{V}_{IV} - \hat{V}_{OLS})^{-1}(\hat{\beta}_{OLS} - \hat{\beta}_{IV}) \rightarrow_d \chi_p^2$$

siendo  $p$  el número de regresores potencialmente endógenos. El contraste de Hausman es un contraste asintótico. Se rechazará la hipótesis nula de exogeneidad cuando  $h > \chi_p^2$ . Si rechazamos la hipótesis nula (existe endogeneidad en el modelo original) tendremos que utilizar el estimador IV, en caso contrario usaremos OLS.

Para el caso de un modelo sobreidentificado, en lugar de comparar los estimadores OLS e IV, tendremos que comparar los estimadores 2SLS e IV repitiendo los mismos pasos.

De manera alternativa, para calcular este contraste también podemos utilizar un procedimiento en dos etapas. Para su ilustración utilizaremos nuestro ejemplo de referencia.

1.ª etapa: Al igual que en 2SLS, hacemos una regresión de  $x_{2i}$  sobre una constante,  $x_{1i}$  y  $z_i$  pero en lugar de calcular el valor ajustado  $\hat{x}_{2i}$  obtenemos los residuos  $\hat{\epsilon}_i$ .

2.ª etapa: Regresamos  $y_i$  sobre el modelo original añadiendo los residuos estimados en la primer etapa:

$$y_i = \pi_0 + \pi_1 x_{1i} + \pi_2 x_{2i} + \pi_3 \hat{\epsilon}_i + v_i$$

Al igual que ocurría con 2SLS, los errores estándar que obtenemos en la segunda etapa no son válidos a menos que todas las variables sean exógenas.

Contrastamos la hipótesis nula de que  $x_{2i}$  es exógena. Bajo exogeneidad, el coeficiente del residuo de la primera etapa no debería ser significativo,  $H_0: \pi_3 = 0$ . Si rechazamos  $H_0$ , encontramos evidencia en contra de la exogeneidad de  $x_{2i}$ , y por tanto en contra de estimar el modelo por OLS. Para poder llevar a cabo este contraste es necesario contar con un instrumento exógeno.

Si en lugar de una variable endógena tuviésemos varias, en la primera etapa calcularíamos los residuos correspondientes a cada regresión, y en la segunda etapa tendríamos que calcular el estadístico  $F$  en lugar del estadístico  $t$ .

## Contrastes sobre la validez de los instrumentos

Si tenemos más instrumentos que variables endógenas, podemos usarlos todos o optar por usar solamente una parte de ellos. Si todos los instrumentos son válidos, al incluirlos en nuestro modelo mejoraríamos la precisión del estimador 2SLS y reduciríamos su error estándar.

## Contraste de relevancia

Supongamos que en el modelo de referencia tenemos varios instrumentos disponibles  $z_{1i}, \dots, z_{mi}$  para el regresor potencialmente endógeno  $x_{2i}$ . Para analizar la relevancia de los instrumentos, estimamos la siguiente ecuación auxiliar:

$$y_i = \beta_0 + \beta_1 x_{1i} + \gamma_1 z_{1i} + \dots + \gamma_m z_{mi} + \epsilon_i$$

Se contrasta la hipótesis nula de que los coeficientes de los instrumentos son conjuntamente iguales a cero:

$$H_0 : \gamma_1 = \dots = \gamma_m = 0$$

El estadístico  $F$  correspondiente da una medida de la información incluida en los instrumentos. Los instrumentos serán débiles (weak instruments) si explican una pequeña proporción de la variación de  $x_{2i}$  (regla práctica sencilla:  $F$  menor que 10).

Si los instrumentos son débiles, los estimadores 2SLS estarán sesgados y cualquier inferencia estadística basada en ellos no será correcta (los estadísticos  $t$  y los intervalos de confianza serán poco fiables). Este problema persiste incluso en muestras grandes.

Si se tienen varios instrumentos, se suele proceder descartando el más débil y con los restantes volver a realizar el análisis hasta encontrar un subconjunto de instrumentos que se puedan considerar relevantes.

## Contraste de sobreidentificación

Si dudamos sobre la exogeneidad de algunos de nuestros instrumentos, incluirlos en el modelo podría causar la inconsistencia del estimador 2SLS. Si el modelo está exactamente identificado, no es posible contrastar la exogeneidad de los instrumentos. Sin embargo, si disponemos de más instrumentos que variables endógenas, es posible realizar un contraste que se llama de sobreidentificación (contraste de Sargan) que resulta extremadamente útil. Se trata de un contraste sobre la exogeneidad de los instrumentos adicionales (sobre los que tenemos dudas) bajo el supuesto de que existen suficientes instrumentos válidos para identificar los parámetros de interés que corresponden a las variables endógenas.

El procedimiento es el siguiente:

- 1.ª etapa: se estima el modelo original via 2SLS y se obtienen los residuos  $\hat{\epsilon}_i$ .
- 2.ª etapa: usando OLS, regresamos los residuos  $\hat{\epsilon}_i$  sobre todas las variables exógenas del modelo y calculamos  $R^2$ . En nuestro ejemplo:

$$\hat{\epsilon}_i = \delta_0 + \delta_1 z_{1i} + \dots + \delta_m z_{mi} + \delta_{m+1} x_{1i} + u_i$$

El estadístico para el contraste es  $NR^2$ , que bajo la hipótesis nula de que los instrumentos son exógenos verifica:

$$LM = NR^2 \rightarrow_d \chi_r^2$$

donde  $r$  es el número de instrumentos adicionales (el grado de sobreidentificación), esto es el número de instrumentos menos el número de regresores endógenos,  $r = q - k$ . En nuestro ejemplo,  $r = m - 1$ .

## Ejemplo 1 (Tasa de retorno de la Educación: Salarios de las mujeres casadas)

La librería **Wooldridge** incluye todos los datos de los ejemplos del manual de econometría Wooldridge (2016).

```
if(!require(wooldridge)){install.packages("wooldridge")}
library(wooldridge)
ls("package:wooldridge")
```

```
## [1] "admnrev"      "affairs"      "airfare"      "alcohol"
## [5] "apple"        "athlet1"      "athlet2"      "attend"
## [9] "audit"        "barium"       "beauty"       "benefits"
## [13] "beveridge"    "big9salary"   "bwght"        "bwght2"
## [17] "campus"       "card"         "cement"       "ceosal1"
## [21] "ceosal2"      "charity"      "consump"      "corn"
## [25] "cps78_85"     "cps91"        "crime1"       "crime2"
## [29] "crime3"       "crime4"       "discrim"      "driving"
## [33] "earns"        "elem94_95"    "engin"        "expendshares"
## [37] "ezanders"     "ezunem"       "fair"         "fertil1"
## [41] "fertil2"      "fertil3"      "fish"         "fringe"
## [45] "gpa1"         "gpa2"         "gpa3"         "happiness"
## [49] "hprice1"      "hprice2"      "hprice3"      "hseinv"
## [53] "htv"          "infmrt"       "injury"       "intdef"
## [57] "intqrt"       "inven"        "jtrain"       "jtrain2"
## [61] "jtrain3"      "k401k"        "k401ksubs"    "kielmc"
## [65] "lawsch85"     "loanapp"      "lowbrth"      "mathpnl"
## [69] "meap00_01"    "meap01"       "meap93"       "minwage"
## [73] "mlb1"         "mroz"         "murder"       "nbasal"
## [77] "nyse"         "okun"         "openness"     "pension"
## [81] "phillips"     "pntsprd"      "prison"       "prminwge"
## [85] "rdchem"       "rdtelec"      "recid"        "rental"
## [89] "return"       "saving"       "sleep75"      "slp75_81"
## [93] "smoke"        "traffic1"     "traffic2"     "twoyear"
## [97] "volat"        "vote1"        "vote2"        "voucher"
## [101] "wage1"        "wage2"        "wagepan"      "wageprc"
## [105] "wine"
```

El capítulo 15 se refiere al tema de variables instrumentales. El Ejemplo 15.1 analiza los rendimientos de la educación para las mujeres casadas utilizando el fichero de datos **mroz**.

```
data("mroz")
str(mroz)
```

```
## 'data.frame':    753 obs. of  22 variables:
## $ inlf      : int  1 1 1 1 1 1 1 1 1 1 ...
## $ hours     : int 1610 1656 1980 456 1568 2032 1440 1020 1458 1600 ...
## $ kidslt6   : int  1 0 1 0 1 0 0 0 0 0 ...
## $ kidsge6   : int  0 2 3 3 2 0 2 0 2 2 ...
## $ age       : int  32 30 35 34 31 54 37 54 48 39 ...
## $ educ      : int  12 12 12 12 14 12 16 12 12 12 ...
## $ wage      : num  3.35 1.39 4.55 1.1 4.59 ...
## $ repwage   : num  2.65 2.65 4.04 3.25 3.6 ...
## $ hushrs    : int  2708 2310 3072 1920 2000 1040 2670 4120 1995 2100 ...
## $ husage    : int  34 30 40 53 32 57 37 53 52 43 ...
## $ huseduc   : int  12 9 12 10 12 11 12 8 4 12 ...
## $ huswage   : num  4.03 8.44 3.58 3.54 10 ...
## $ faminc    : num 16310 21800 21040 7300 27300 ...
## $ mtr       : num  0.721 0.661 0.692 0.781 0.622 ...
## $ motheduc  : int  12 7 12 7 12 14 14 3 7 7 ...
## $ fatheduc  : int  7 7 7 7 14 7 7 3 7 7 ...
## $ unem      : num  5 11 5 5 9.5 7.5 5 5 3 5 ...
## $ city      : int  0 1 0 0 1 1 0 0 0 0 ...
## $ exper     : int  14 5 15 6 7 33 11 35 24 21 ...
## $ nwifeinc  : num  10.9 19.5 12 6.8 20.1 ...
## $ lwage     : num  1.2102 0.3285 1.5141 0.0921 1.5243 ...
## $ expersq   : int  196 25 225 36 49 1089 121 1225 576 441 ...
## - attr(*, "datalabel")= chr ""
## - attr(*, "time.stamp")= chr "25 Jun 2011 23:03"
## - attr(*, "formats")= chr  "%9.0g" "%9.0g" "%9.0g" "%9.0g" ...
## - attr(*, "types")= int   251 252 251 251 251 251 254 254 252 251 ...
## - attr(*, "val.labels")= chr  "" "" "" "" ...
## - attr(*, "var.labels")= chr  "=1 if in lab frce, 1975" "hours worked, 1975" "# kids < 6 years"
## kids 6-18" ...
## - attr(*, "version")= int 10
```

```
names(mroz)
```

```
## [1] "inlf"      "hours"     "kidslt6"   "kidsge6"   "age"       "educ"
## [7] "wage"      "repwage"   "hushrs"    "husage"     "huseduc"   "huswage"
## [13] "faminc"    "mtr"       "motheduc"  "fatheduc"   "unem"      "city"
## [19] "exper"     "nwifeinc"  "lwage"     "expersq"
```

```
attributes(mroz)$var.labels
```

```
## [1] "=1 if in lab frce, 1975"      "hours worked, 1975"
## [3] "# kids < 6 years"             "# kids 6-18"
## [5] "woman's age in yrs"           "years of schooling"
## [7] "est. wage from earn, hrs"     "rep. wage at interview in 1976"
## [9] "hours worked by husband, 1975" "husband's age"
## [11] "husband's years of schooling" "husband's hourly wage, 1975"
## [13] "family income, 1975"          "fed. marg. tax rte facing woman"
## [15] "mother's years of schooling"  "father's years of schooling"
## [17] "unem. rate in county of resid." "=1 if live in SMSA"
## [19] "actual labor mkt exper"       "(faminc - wage*hours)/1000"
## [21] "log(wage)"                    "exper^2"
```

Realizamos una rápida inspección de los datos. ¿Por qué el salario medio de las mujeres que no participan en el mercado laboral es *NaN*?

```
table(mroz$inlf)
```

```
##
## 0 1
## 325 428
```

```
if(!require(dplyr)){install.packages("dplyr")}
library(dplyr)
mroz%>%
  group_by(inlf)%>%
  summarise(mean(lwage, na.rm = TRUE))
```

```
## # A tibble: 2 x 2
##   inlf `mean(lwage, na.rm = TRUE)`
##   <int>           <dbl>
## 1     0           NaN
## 2     1           1.19
```

Exploramos si nuestra base de datos contiene *missing values* en las demás variables.

```
has_na <- sapply(mroz, function(x) any(is.na(x)))
has_na
```

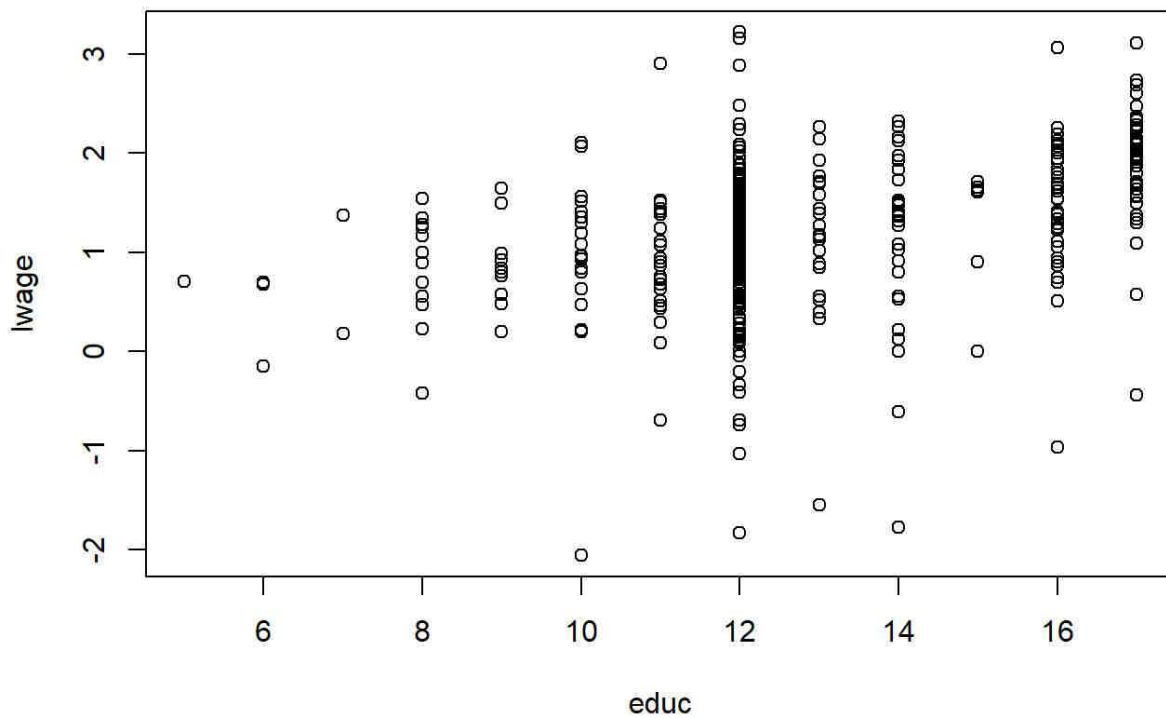
```
##   inlf   hours kidslt6 kidsge6   age   educ   wage repwage
## FALSE FALSE   FALSE   FALSE FALSE FALSE   TRUE   FALSE
## hushrs  husage huseduc huswage faminc   mtr motheduc fatheduc
## FALSE FALSE   FALSE   FALSE FALSE FALSE   FALSE   FALSE
##   unem   city   exper nwifeinc   lwage expersq
## FALSE FALSE   FALSE   FALSE   TRUE  FALSE
```

```
count_na <- sapply(mroz, function(x) sum(is.na(x)))
count_na
```

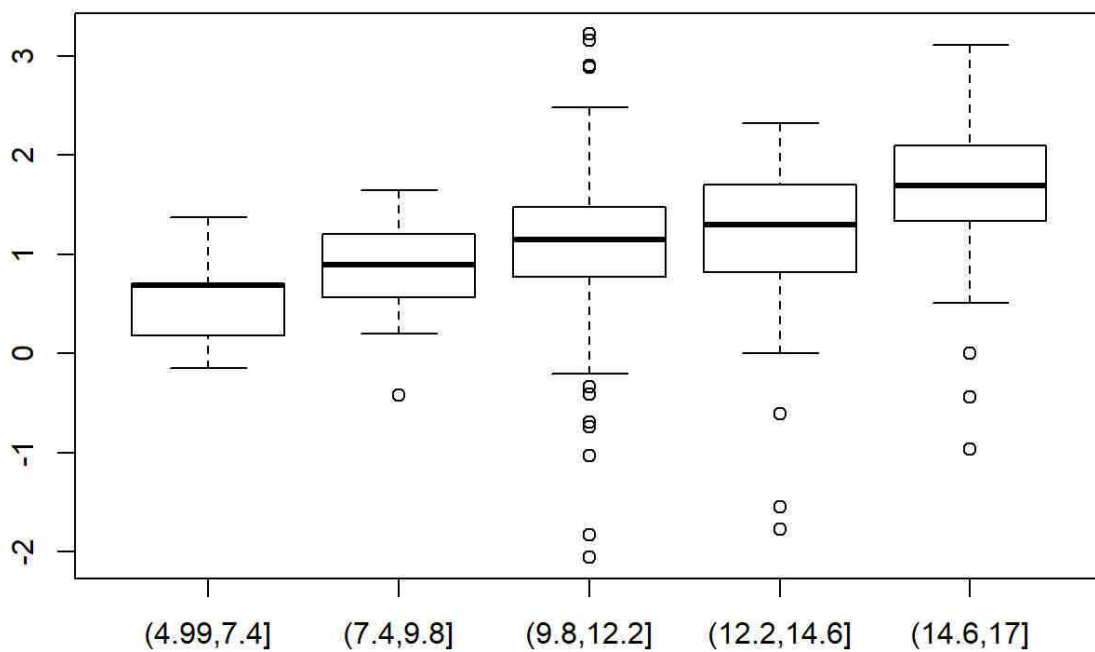
```
##   inlf   hours kidslt6 kidsge6   age   educ   wage repwage
##    0     0     0     0     0     0    325     0
## hushrs  husage huseduc huswage faminc   mtr motheduc fatheduc
##    0     0     0     0     0     0     0     0
##   unem   city   exper nwifeinc   lwage expersq
##    0     0     0     0    325     0
```

¿Cómo espera que evolucione el salario medio a medida que aumentan los años de estudio?

```
plot(lwage ~ educ, data = mroz)
```



```
boxplot(lwage~cut(educ, breaks = 5), data=mroz)
```



En primer lugar, calculamos el estimador OLS utilizando el comando `lm`.

```
modelo.lm <- lm( lwage ~ educ, data = mroz, subset = (inlf==1))
summary(modelo.lm)
```

```
##
## Call:
## lm(formula = lwage ~ educ, data = mroz, subset = (inlf == 1))
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -3.10256 -0.31473  0.06434  0.40081  2.10029
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  -0.1852     0.1852  -1.000   0.318
## educ          0.1086     0.0144   7.545 2.76e-13 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.68 on 426 degrees of freedom
## Multiple R-squared:  0.1179, Adjusted R-squared:  0.1158
## F-statistic: 56.93 on 1 and 426 DF,  p-value: 2.761e-13
```

También podríamos haber seleccionado previamente la submuestra de mujeres que participan en el mercado laboral.

```
misdatos1 <- subset(mroz,!is.na(wage))
modelo1.lm <- lm( lwage ~ educ , data = misdatos1)
summary(modelo1.lm)
```

```
##
## Call:
## lm(formula = lwage ~ educ, data = misdatos1)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -3.10256 -0.31473  0.06434  0.40081  2.10029
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  -0.1852     0.1852  -1.000   0.318
## educ          0.1086     0.0144   7.545 2.76e-13 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.68 on 426 degrees of freedom
## Multiple R-squared:  0.1179, Adjusted R-squared:  0.1158
## F-statistic: 56.93 on 1 and 426 DF,  p-value: 2.761e-13
```

```
misdatos2 <- mroz[complete.cases(mroz),]
modelo2.lm <- lm( lwage ~ educ, data = misdatos2)
summary(modelo2.lm)
```

```
##
## Call:
## lm(formula = lwage ~ educ, data = misdatos2)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -3.10256 -0.31473  0.06434  0.40081  2.10029
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  -0.1852     0.1852  -1.000   0.318
## educ           0.1086     0.0144   7.545 2.76e-13 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.68 on 426 degrees of freedom
## Multiple R-squared:  0.1179, Adjusted R-squared:  0.1158
## F-statistic: 56.93 on 1 and 426 DF,  p-value: 2.761e-13
```

```
with(misdatos1,summary(misdatos1[,-18]))
```



```
##      inlf      hours      kidslt6      kidsge6
## Min.   :1   Min.   : 12.0   Min.   :0.0000   Min.   :0.00
## 1st Qu.:1   1st Qu.: 609.5   1st Qu.:0.0000   1st Qu.:0.00
## Median :1   Median :1365.5   Median :0.0000   Median :1.00
## Mean   :1   Mean   :1302.9   Mean   :0.1402   Mean   :1.35
## 3rd Qu.:1   3rd Qu.:1910.5   3rd Qu.:0.0000   3rd Qu.:2.00
## Max.   :1   Max.   :4950.0   Max.   :2.0000   Max.   :8.00
##      age      educ      wage      repwage
## Min.   :30.00   Min.   : 5.00   Min.   : 0.1282   Min.   :0.000
## 1st Qu.:35.00   1st Qu.:12.00   1st Qu.: 2.2626   1st Qu.:1.420
## Median :42.00   Median :12.00   Median : 3.4819   Median :3.195
## Mean   :41.97   Mean   :12.66   Mean   : 4.1777   Mean   :3.186
## 3rd Qu.:47.25   3rd Qu.:14.00   3rd Qu.: 4.9707   3rd Qu.:4.550
## Max.   :60.00   Max.   :17.00   Max.   :25.0000   Max.   :9.980
##      hushrs      husage      huseduc      huswage
## Min.   : 175   Min.   :30.00   Min.   : 4.00   Min.   : 0.5128
## 1st Qu.:1920   1st Qu.:38.00   1st Qu.:11.00   1st Qu.: 4.8217
## Median :2106   Median :45.00   Median :12.00   Median : 6.6831
## Mean   :2233   Mean   :44.61   Mean   :12.61   Mean   : 7.2262
## 3rd Qu.:2504   3rd Qu.:51.00   3rd Qu.:16.00   3rd Qu.: 8.8378
## Max.   :5010   Max.   :60.00   Max.   :17.00   Max.   :26.5780
##      faminc      mtr      motheduc      fatheduc
## Min.   : 2400   Min.   :0.4415   Min.   : 0.000   Min.   : 0.000
## 1st Qu.:16286   1st Qu.:0.6215   1st Qu.: 7.000   1st Qu.: 7.000
## Median :21961   Median :0.6915   Median :10.000   Median : 7.000
## Mean   :24130   Mean   :0.6683   Mean   : 9.516   Mean   : 8.988
## 3rd Qu.:29793   3rd Qu.:0.7215   3rd Qu.:12.000   3rd Qu.:12.000
## Max.   :91044   Max.   :0.9415   Max.   :17.000   Max.   :17.000
##      unem      exper      nwifeinc      lwage
## Min.   : 3.000   Min.   : 0.00   Min.   : -0.02906   Min.   : -2.0542
## 1st Qu.: 7.500   1st Qu.: 7.00   1st Qu.:12.36525   1st Qu.: 0.8165
## Median : 7.500   Median :12.00   Median :17.08000   Median : 1.2476
## Mean   : 8.546   Mean   :13.04   Mean   :18.93748   Mean   : 1.1902
## 3rd Qu.:11.000   3rd Qu.:18.00   3rd Qu.:23.51500   3rd Qu.: 1.6036
## Max.   :14.000   Max.   :38.00   Max.   :91.00000   Max.   : 3.2189
##      expersq
## Min.   : 0.0
## 1st Qu.: 49.0
## Median :144.0
## Mean   :234.7
## 3rd Qu.:324.0
## Max.   :1444.0
```

```
#with(misdatos1,split(misdatos1,misdatos1))
```

El paquete **AER** contiene el comando `ivreg` que permite calcular el estimador de IV. Su funcionamiento es similar al del comando `lm` utilizado para calcular el estimador OLS.

```
if(!require(AER)){install.packages("AER")}
library(AER)
ls("package:AER")
```

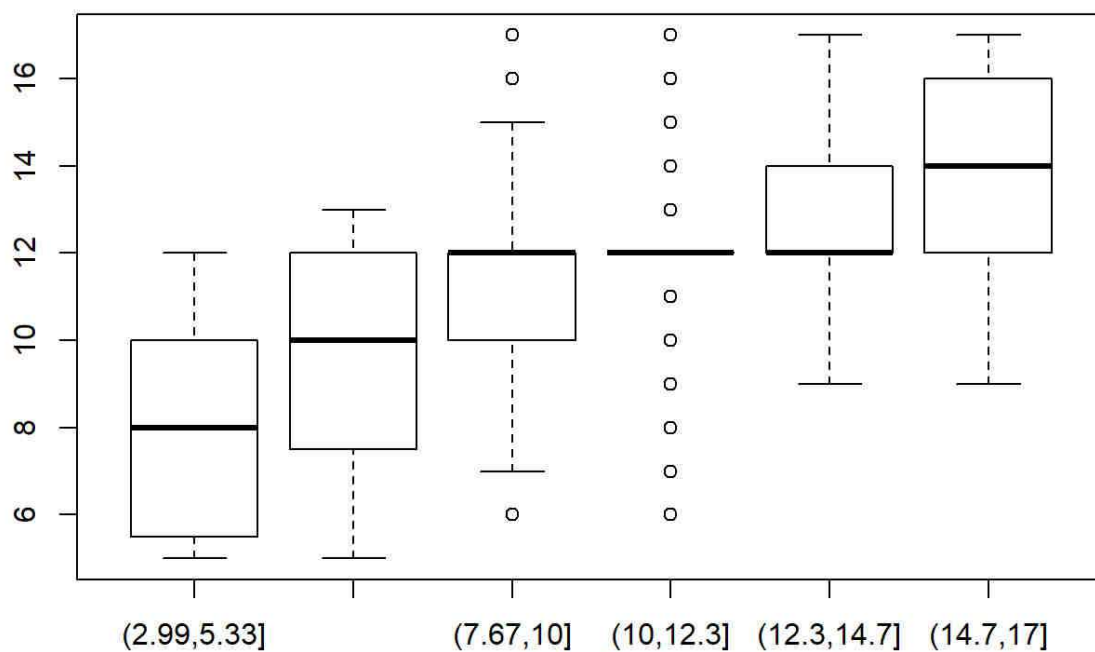
```
## [1] "dispersiontest" "ivreg"          "ivreg.fit"      "tobit"
```

Sospechamos que la variable *educ* es endógena y buscamos candidatos para instrumentar el nivel educativo. Los instrumentos...

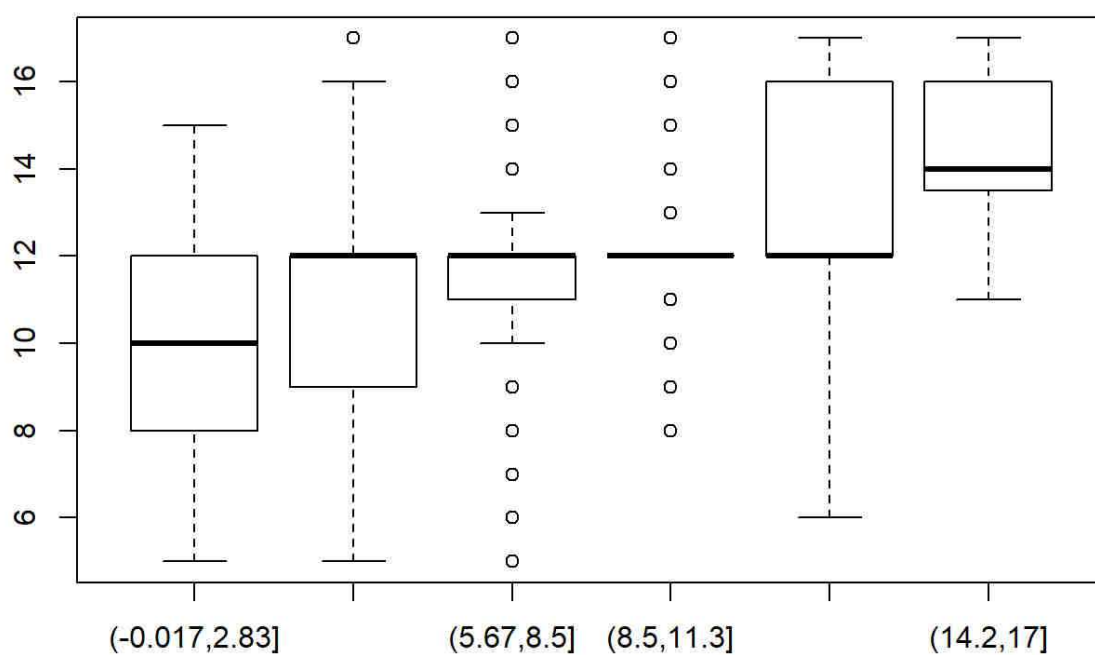
1. ...tienen que estar correlacionados con la variable endógena
2. ...pero NO con la variable omitida (incluida en el término de error)

En la base de datos hay tres posibles candidatos:

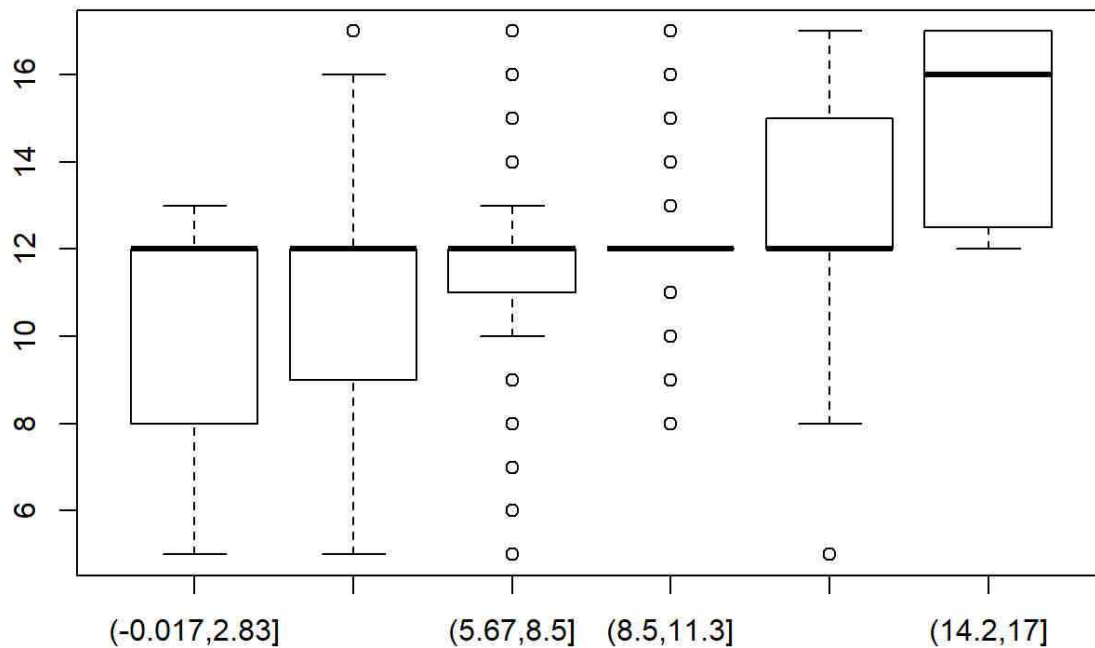
```
boxplot(educ ~ cut(huseduc, breaks = 6), data=mroz)
```



```
boxplot(educ ~ cut(motheduc, breaks = 6), data=mroz)
```



```
boxplot(educ ~ cut(fatheduc, breaks = 6), data=mroz)
```



Comprobamos si se cumple el punto 1:

```
cor.test(mroz$educ, mroz$huseduc)
```

```
##
## Pearson's product-moment correlation
##
## data: mroz$educ and mroz$huseduc
## t = 21.204, df = 751, p-value < 2.2e-16
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
##  0.5652203 0.6547720
## sample estimates:
##          cor
## 0.6119538
```

```
cor.test(mroz$educ, mroz$motheduc)
```

```
##
## Pearson's product-moment correlation
##
## data: mroz$educ and mroz$motheduc
## t = 13.252, df = 751, p-value < 2.2e-16
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
##  0.3755721 0.4914954
## sample estimates:
##          cor
## 0.4353365
```

```
cor.test(mroz$educ, mroz$fatheduc)
```

```
##
## Pearson's product-moment correlation
##
## data: mroz$educ and mroz$fatheduc
## t = 13.521, df = 751, p-value < 2.2e-16
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
##  0.3831236 0.4981564
## sample estimates:
##          cor
## 0.4424582
```

Calculamos el estimador IV (mínimos cuadrados en dos etapas) de modo automático con el comando `ivreg` :

```
modelo.iv <- ivreg(lwage ~ educ | fatheduc + huseduc + motheduc, data = mroz, subset = (inlf==1))
summary(modelo.iv, diagnostics=TRUE)
```

```
##
## Call:
## ivreg(formula = lwage ~ educ | fatheduc + huseduc + motheduc,
##       data = mroz, subset = (inlf == 1))
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -3.05569 -0.33077  0.06042  0.41166  2.07726
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  0.25721     0.28381   0.906  0.36530
## educ         0.07370     0.02227   3.310  0.00101 **
##
## Diagnostic tests:
##              df1 df2 statistic p-value
## Weak instruments    3 424   104.036 <2e-16 ***
## Wu-Hausman          1 425    4.370  0.0372 *
## Sargan              2  NA     1.416  0.4926
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.6847 on 426 degrees of freedom
## Multiple R-Squared: 0.1057, Adjusted R-squared: 0.1036
## Wald test: 10.96 on 1 and 426 DF, p-value: 0.001013
```

Si comparamos el estimador OLS y el estimador de IV de la variable *educ* vemos que el estimador OLS sobreestima (sesgo positivo) el efecto de la educación.

```
if(!require(stargazer)){install.packages("stargazer")}
library(stargazer)
stargazer(modelo.lm,modelo.iv,type="text",keep.stat=c("n","rsq"))
```

```
##
## =====
##               Dependent variable:
##               -----
##               lwage
##               OLS      instrumental
##               variable
##               (1)      (2)
## -----
## educ          0.109***    0.074***
##               (0.014)    (0.022)
##
## Constant      -0.185      0.257
##               (0.185)    (0.284)
##
## -----
## Observations   428        428
## R2             0.118      0.106
## =====
## Note:          *p<0.1; **p<0.05; ***p<0.01
```

Si sospechamos que nuestro modelo puede presentar heterocedasticidad, realizamos el contraste de Breusch-Pagan. La hipótesis nula es  $H_0$  : *Homocedasticidad*. La librería **lmtest** incluye el comando `bptest` que permite realizar este contraste pero solo sobre las variables originales del modelo (no las transformadas)

```
if(!require(lmtest)){install.packages("lmtest")}
library(lmtest)
lmtest::bptest(modelo.lm)
```

```
##
## studentized Breusch-Pagan test
##
## data:  modelo.lm
## BP = 0.30859, df = 1, p-value = 0.5785
```

Como rechazamos la hipótesis nula de homocedasticidad, necesitamos un estimador robusto de la matriz de varianzas-covarianzas. Para calcular errores estándar robustos a heterocedasticidad podemos utilizar los comandos `hcc` (paquete **car**) o `vcovHC` (paquete **sandwich**). Si queremos obtener el estimador robusto de White (1980):

```
if(!require(sandwich)){install.packages("sandwich")}
library(sandwich)
lmtest::coeftest(modelo.lm, vcov=sandwich)
```

```
##
## t test of coefficients:
##
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept) -0.185197   0.170349  -1.0872   0.2776
## educ         0.108649   0.013384   8.1178 5.149e-15 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
lmtest::coeftest(modelo.lm, vcov=vcovHC, type="HC0")
```

```
##
## t test of coefficients:
##
##           Estimate Std. Error t value Pr(>|t|)
## (Intercept) -0.185197  0.170349 -1.0872  0.2776
## educ        0.108649  0.013384  8.1178 5.149e-15 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Davidson & MacKinnon (1993) recomiendan usar “HC3” en lugar de “HC0”. De hecho, éste es el valor prefijado de `vcovHC` .

```
lmtest::coeftest(modelo.lm, vcov=vcovHC)
```

```
##
## t test of coefficients:
##
##           Estimate Std. Error t value Pr(>|t|)
## (Intercept) -0.185197  0.171834 -1.0778  0.2817
## educ        0.108649  0.013506  8.0447 8.665e-15 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
lmtest::coeftest(modelo.lm, vcov=vcovHC, type="HC3")
```

```
##
## t test of coefficients:
##
##           Estimate Std. Error t value Pr(>|t|)
## (Intercept) -0.185197  0.171834 -1.0778  0.2817
## educ        0.108649  0.013506  8.0447 8.665e-15 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Si sospechamos que hay heterocedasticidad, calculamos los errores estándar robustos:

```
summary(modelo.iv, vcov=sandwich, diagnostics=TRUE)
```

```
##
## Call:
## ivreg(formula = lwage ~ educ | fatheduc + huseduc + motheduc,
##       data = mroz, subset = (inlf == 1))
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -3.05569 -0.33077  0.06042  0.41166  2.07726
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)   0.2572     0.2817   0.913  0.36176
## educ          0.0737     0.0224   3.291  0.00108 **
##
## Diagnostic tests:
##              df1 df2 statistic p-value
## Weak instruments    3 424   102.264 <2e-16 ***
## Wu-Hausman          1 425    5.181  0.0233 *
## Sargan              2  NA    1.416  0.4926
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.6847 on 426 degrees of freedom
## Multiple R-Squared: 0.1057, Adjusted R-squared: 0.1036
## Wald test: 10.83 on 1 and 426 DF, p-value: 0.001082
```

A continuación realizamos la diagnosis de nuestro modelo. En primer lugar tenemos que determinar si tenemos un problema de variables endógenas. Atendiendo al contraste de Wu-Hausman, con un  $p - value = 0.0233$  rechazamos la hipótesis nula de exogeneidad al 5%, pero no al 1%. Por tanto, si establecemos un nivel de confianza del 5%, tendremos que utilizar el estimador IV porque el estimador OLS no será consistente.

Por último, tenemos que analizar la validez de nuestros instrumentos. El contraste de relevancia (weak instruments) con un  $p - value < 2e - 16$  nos indica que rechazamos  $H_0$  y que, por tanto, los tres instrumentos utilizados son relevantes. Además, el contraste de Sargan arroja un  $p - value = 0.4926$ , esto es, aceptamos la hipótesis nula de que los tres instrumentos son exógenos.

## Referencias

Heiss, F. (2016): Using R for Introductory Econometrics. Create Space. <http://www.urfie.net/read.html> (<http://www.urfie.net/read.html>)

Wooldridge, Jeffrey M. (2016), Introductory Econometrics: A Modern Approach. 6th ed. Cengage Learning, Boston, MA.



Proyecto de innovación  
“Innova Docencia” de la UCM  
Convocatoria 2018-2019  
Proyecto número 275

**R-adaptación de la asignatura de métodos econométricos en economía y finanzas del  
Grado de Estadística Aplicada**

Facultad de Estudios Estadísticos  
Departamento de Economía Aplicada, Pública y Política

**Título**

***Práctica 16: Modelo de datos de Panel con R: la librería “pml”***

**Correspondiente al**

**Capítulo 8 MODELOS DE REGRESIÓN CON DATOS DE PANEL**

- Especificación de modelos con datos de panel
  - Estimación de modelos con efectos fijos
  - Estimación de modelos con efectos aleatorios
  - Sistema de ecuaciones aparentemente no relacionadas
  - Estimador de efectos fijos observables con el estimador CRE
- Práctica 16: Modelo de datos de Panel con R: la librería “pml”***  
***Práctica 17: La ecuación de Salarios con Datos de Panel***



# Modelos de regresión con Datos de Panel en R

## Asignatura Métodos Econométricos en Economía y Finanzas

Profesores  
Loren Escot, Alicia Pérez-Alonso y Julio E. Sandubete

Dpto Economía Aplicada, Pública y Política  
Facultad Estudios Estadísticos - Universidad Complutense de Madrid

Curso académico 2018/2019

### Abstract

En esta práctica se revisa la estimación de modelos de regresión con Datos de Panel. Se presentará la librería `plm` para trabajar con datos de panel, esto es, con datos que tienen una dimensión temporal y otra de sección cruzada. Se presentarán los diferentes métodos de regresión como el modelo pooling, between, Fixed effects y random effects, y se presentarán los diferentes contrastes para poder saber cual de ellos es mejor. También presentamos la estimación CRE de efectos fijos para estimar efectos fijos si observables.

Se presentarán las diferentes técnicas utilizando un ejemplo de líneas aéreas utilizado en el manual de Wooldridge

- 1 Carga de librerías necesarias `plm`
- 2 Carga de datos
- 3 Definiendo la estructura de los datos de panel con `plm`
- 4 Estadística descriptiva del panel de datos
- 5 Métodos de estimación del modelo lineal con datos de panel
  - 5.1 Pooled OLS estimator
- 6 Between estimator
- 7 Estimador cuando hay un problema de endogeneidad con los efectos fijos no observables
  - 7.1 Estimador de Primeras diferencias y de Efectos fijos
  - 7.2 Estimador de efectos Fijos
- 8 Comparación de modelos
- 9 Efectos temporales
- 10 Estimador de Efectos Aleatorios
- 11 DIAGNOSIS de los modelos: ¿qué método de regresión debería utilizar?
  - 11.1 ¿Existen efectos individuales? (paneles heterogéneos)
  - 11.2 ¿Tengo un problema de endogeneidad?
- 12 El problema de la heterocedasticidad entre individuos
  - 12.1 Estimación Robusta de las Varianzas de los estimadores
  - 12.2 Estimación de efectos fijos por mínimos Cuadrados Generalizados (Cluster de individuos)
- 13 ¿Estimador CRE Se podría obtener de alguna manera la estimación de los efectos fijos individuales si observables?

```
#=====
```

```
# Datos de panel en R
```

# 1 Carga de librerías necesarias plm

```
suppressMessages(library(stargazer))
suppressMessages(library(wooldridge)) #para utilizar los datos de ejemplo del libro de Wooldridge
suppressMessages(library(plm)) # librería para estimar con datos de panel
```

## 2 Carga de datos

Trabajaremos con los datos de líneas aéreas:

Construiremos un modelo donde la la tarifa de billete solo ida (**fare**), dependerá de la distancia del vuelo (**dist**) (*a mayor distancia mayor coste y mayor precio*) y de la concentraciónn del mercado en pocas líneas (**concen**) (*a mayor competencia menor precio*)

```
data("airfare")
# para obtenr una ayuda del contenido de los datos
# ?airfare

# trabajaremos con los datos de líneas aéreas:
# construiremos un modelo donde la la tarifa de billete solo ida (fare),
# dependerá de la distancia del vuelo (dist) (a mayor distancia mayor coste y mayor precio)
# y de la concentraciónn del mercado en pocas líneas (concen) (a mayor competencia menor precio)

summary(airfare)
```

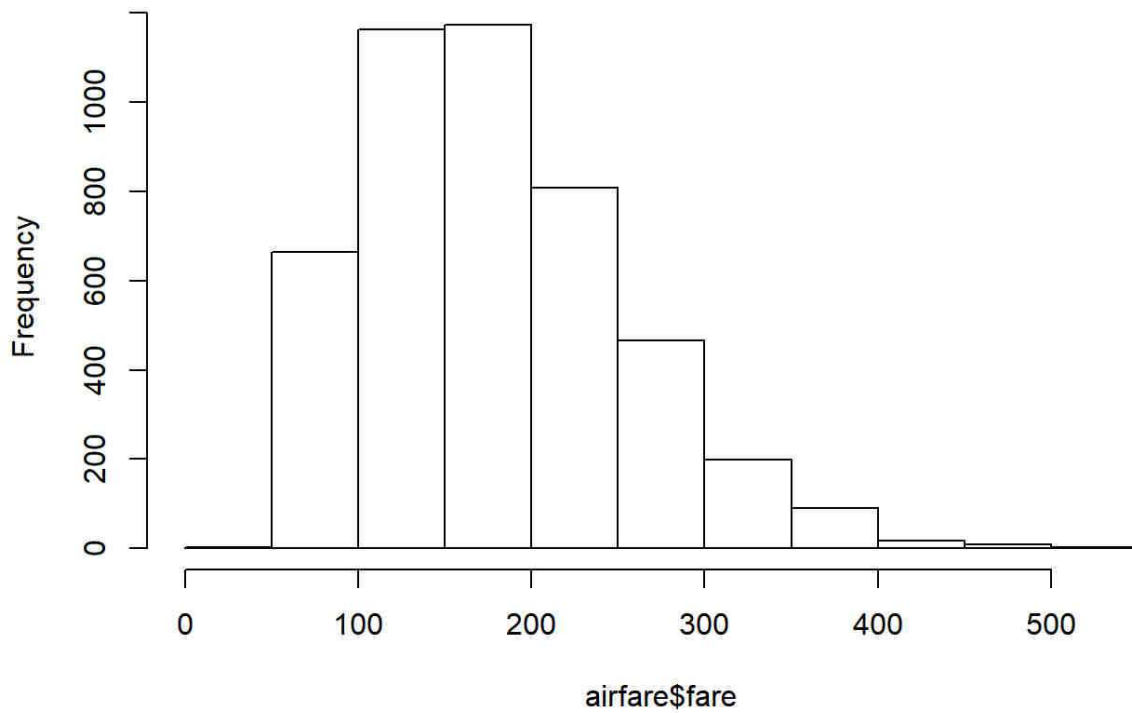
```
##      year      id      dist      passen
## Min.   :1997   Min.    :  1   Min.    : 95.0   Min.    :  2.0
## 1st Qu.:1998   1st Qu.: 288   1st Qu.: 505.0   1st Qu.: 215.0
## Median :1998   Median : 575   Median : 861.0   Median : 357.0
## Mean   :1998   Mean    : 575   Mean    : 989.7   Mean    : 636.8
## 3rd Qu.:1999   3rd Qu.: 862   3rd Qu.:1304.0   3rd Qu.: 717.0
## Max.   :2000   Max.    :1149   Max.    :2724.0   Max.    :8497.0
##      fare      bmktsshr      ldist      y98
## Min.    : 37.0   Min.    :0.1605   Min.    :4.554   Min.    :0.00
## 1st Qu.:123.0   1st Qu.:0.4650   1st Qu.:6.225   1st Qu.:0.00
## Median :168.0   Median :0.6039   Median :6.758   Median :0.00
## Mean    :178.8   Mean     :0.6101   Mean     :6.696   Mean     :0.25
## 3rd Qu.:225.0   3rd Qu.:0.7531   3rd Qu.:7.173   3rd Qu.:0.25
## Max.    :522.0   Max.     :1.0000   Max.     :7.910   Max.     :1.00
##      y99      y00      lfare      ldistsq
## Min.    :0.00   Min.    :0.00   Min.    :3.611   Min.    :20.74
## 1st Qu.:0.00   1st Qu.:0.00   1st Qu.:4.812   1st Qu.:38.75
## Median :0.00   Median :0.00   Median :5.124   Median :45.67
## Mean    :0.25   Mean     :0.25   Mean     :5.096   Mean     :45.28
## 3rd Qu.:0.25   3rd Qu.:0.25   3rd Qu.:5.416   3rd Qu.:51.45
## Max.    :1.00   Max.     :1.00   Max.     :6.258   Max.     :62.57
##      concen      lpassen
## Min.    :0.1605   Min.    :0.6931
## 1st Qu.:0.4650   1st Qu.:5.3706
## Median :0.6039   Median :5.8777
## Mean    :0.6101   Mean     :6.0170
## 3rd Qu.:0.7531   3rd Qu.:6.5751
## Max.    :1.0000   Max.     :9.0475
```

```
head(airfare,n = 20)
```

##	year	id	dist	passen	fare	bmktshr	ldist	y98	y99	y00	lfare	ldistsq
## 1	1997	1	528	152	106	0.8386	6.269096	0	0	0	4.663439	39.30157
## 2	1998	1	528	265	106	0.8133	6.269096	1	0	0	4.663439	39.30157
## 3	1999	1	528	336	113	0.8262	6.269096	0	1	0	4.727388	39.30157
## 4	2000	1	528	298	123	0.8612	6.269096	0	0	1	4.812184	39.30157
## 5	1997	2	861	282	104	0.5798	6.758094	0	0	0	4.644391	45.67184
## 6	1998	2	861	178	105	0.5817	6.758094	1	0	0	4.653960	45.67184
## 7	1999	2	861	204	115	0.7319	6.758094	0	1	0	4.744932	45.67184
## 8	2000	2	861	190	129	0.5386	6.758094	0	0	1	4.859812	45.67184
## 9	1997	3	852	241	207	0.8180	6.747587	0	0	0	5.332719	45.52993
## 10	1998	3	852	253	188	0.8172	6.747587	1	0	0	5.236442	45.52993
## 11	1999	3	852	244	229	0.7998	6.747587	0	1	0	5.433722	45.52993
## 12	2000	3	852	244	247	0.7097	6.747587	0	0	1	5.509388	45.52993
## 13	1997	4	724	233	243	0.4604	6.584791	0	0	0	5.493062	43.35947
## 14	1998	4	724	248	226	0.4614	6.584791	1	0	0	5.420535	43.35947
## 15	1999	4	724	224	229	0.4334	6.584791	0	1	0	5.433722	43.35947
## 16	2000	4	724	339	176	0.3716	6.584791	0	0	1	5.170484	43.35947
## 17	1997	5	1073	396	119	0.4571	6.978214	0	0	0	4.779123	48.69547
## 18	1998	5	1073	529	105	0.5632	6.978214	1	0	0	4.653960	48.69547
## 19	1999	5	1073	494	121	0.5008	6.978214	0	1	0	4.795791	48.69547
## 20	2000	5	1073	710	109	0.5197	6.978214	0	0	1	4.691348	48.69547
##	concen	lpassen										
## 1	0.8386	5.023880										
## 2	0.8133	5.579730										
## 3	0.8262	5.817111										
## 4	0.8612	5.697093										
## 5	0.5798	5.641907										
## 6	0.5817	5.181784										
## 7	0.7319	5.318120										
## 8	0.5386	5.247024										
## 9	0.8180	5.484797										
## 10	0.8172	5.533390										
## 11	0.7998	5.497168										
## 12	0.7097	5.497168										
## 13	0.4604	5.451038										
## 14	0.4614	5.513429										
## 15	0.4334	5.411646										
## 16	0.3716	5.826000										
## 17	0.4571	5.981414										
## 18	0.5632	6.270988										
## 19	0.5008	6.202536										
## 20	0.5197	6.565265										

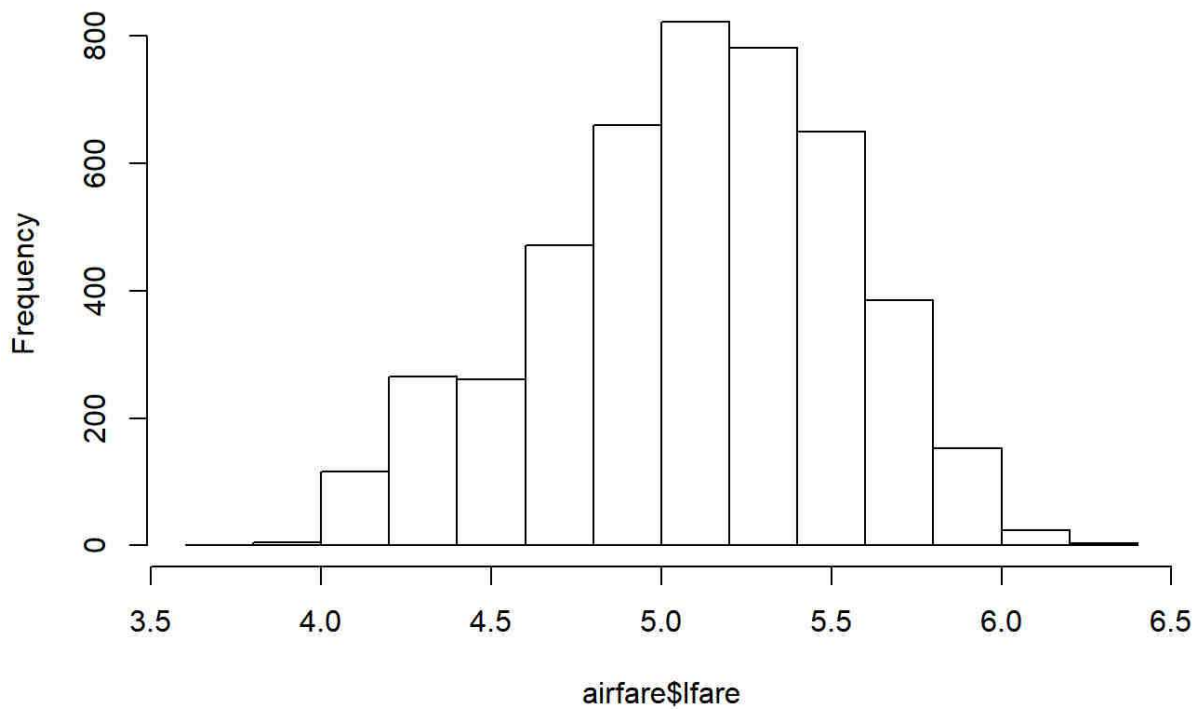
```
#Trabajaremos con las series en logaritmos
hist(airfare$fare)
```

**Histogram of airfare\$fare**



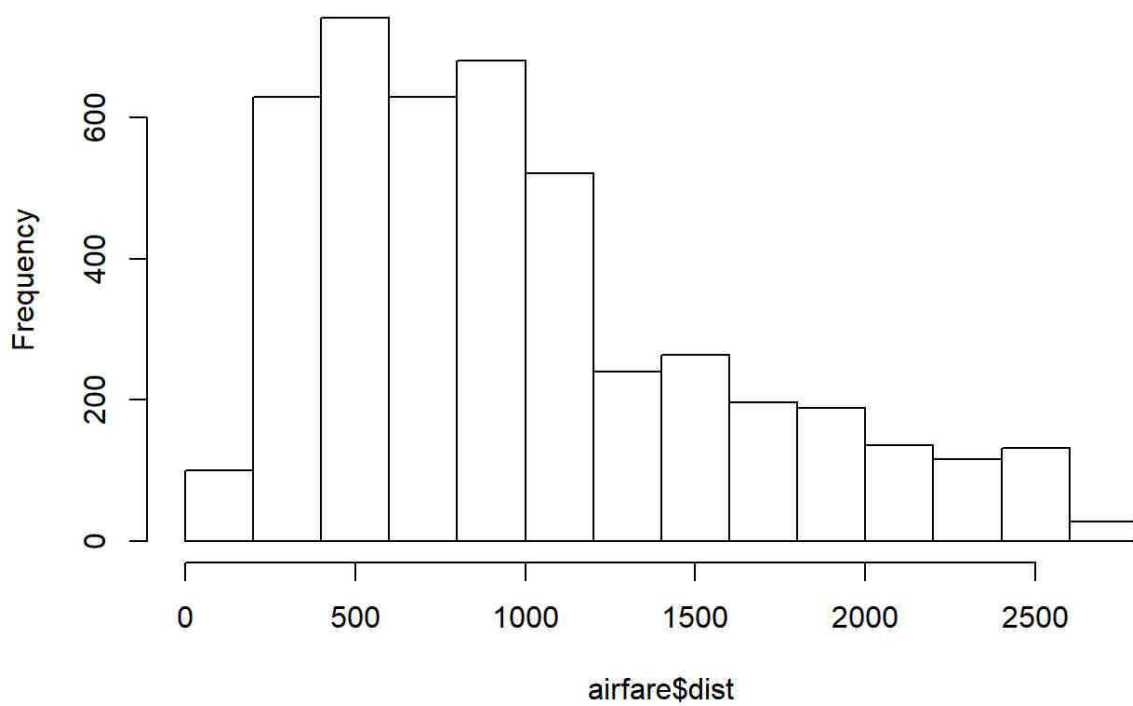
```
hist(airfare$fare)
```

**Histogram of airfare\$lfare**



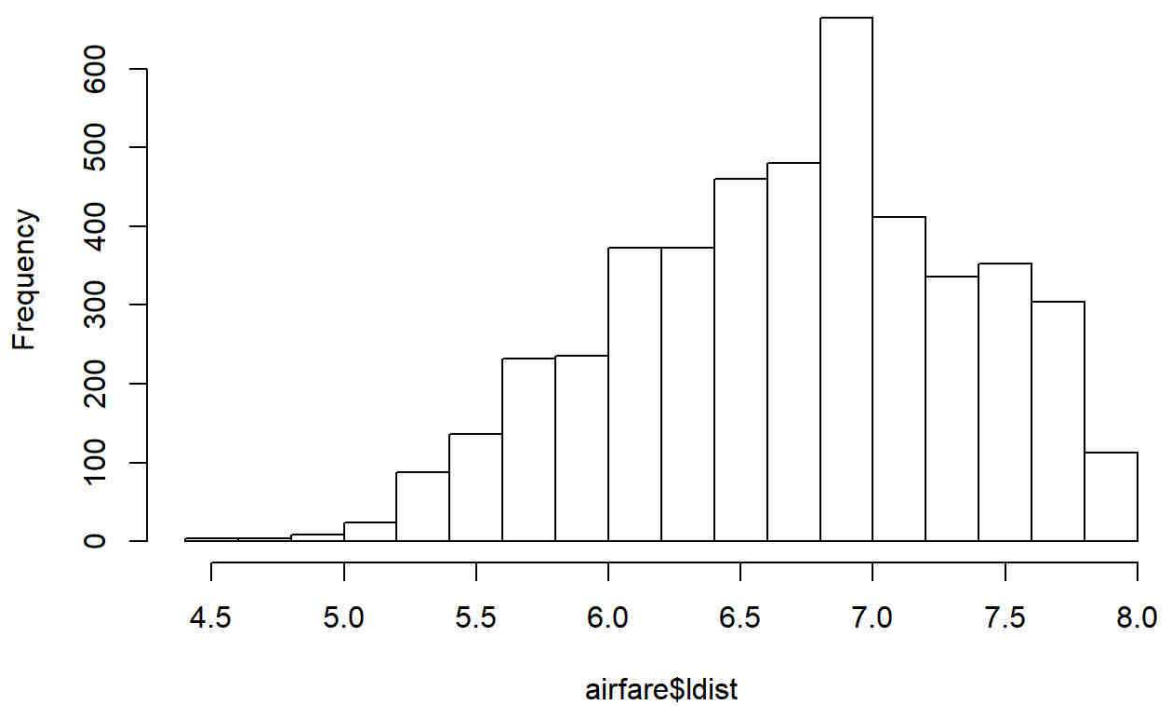
```
hist(airfare$dist)
```

**Histogram of airfare\$ldist**

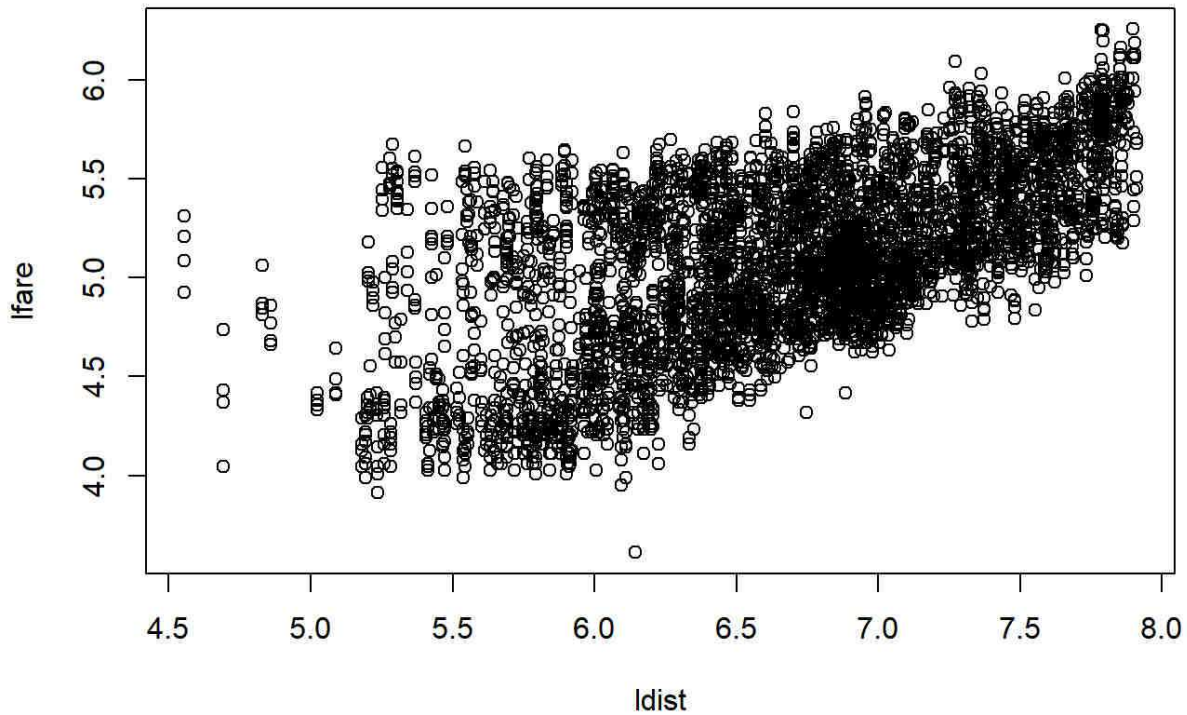


```
hist(airfare$ldist)
```

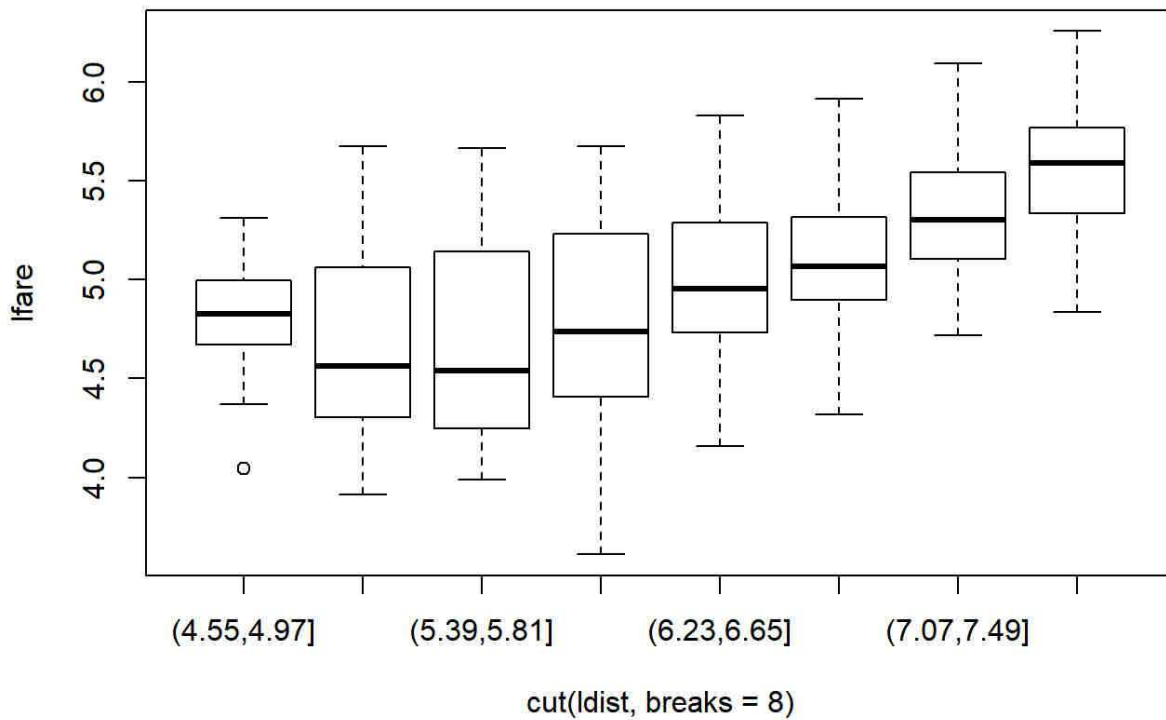
**Histogram of airfare\$ldist**



```
plot(lfare ~ ldist, data=airfare)
```



```
boxplot(lfare~cut(ldist,breaks = 8),data=airfare)
```



3 Definiendo la estructura de los datos de panel con `pm1`

Establecemos los dos índices (de tiempo y de cross-section) para que el conjunto de datos sea un panel de datos (objeto `pdata.frame`)

Vamos a utilizar el "id" de vuelo para el identificador de sección cruzada y "year" como identificador del tiempo

```
# Definimos la estructura de los datos de Panel
# install.packages("plm")
# library(plm) # librería para estimar con datos de panel

# Establecemos los dos índices (de tiempo y de cross-section) para
# que el conjunto de datos sea un panel de datos (objeto pdata.frame)

# Vamos a utilizar el "id" de vuelo para el identificador de sección cruzada
# y "year" como identificador del tiempo

# tenemos 4 años
table(airfare$year)
```

```
##
## 1997 1998 1999 2000
## 1149 1149 1149 1149
```

```
# y 1149 vuelos
nrow(table(airfare$id))
```

```
## [1] 1149
```

```
# aquí definimos la estructura del panel en un nuevo objeto `plm.data`

# Creación del panel por defecto
pdata <- pdata.frame(airfare, index=c("id", "year"))
str(pdata)
```

```

## Classes 'pdata.frame' and 'data.frame': 4596 obs. of 14 variables:
## $ year : Factor w/ 4 levels "1997","1998",...: 1 2 3 4 1 2 3 4 1 2 ...
## .. attr(*, "names")= chr "1-1997" "1-1998" "1-1999" "1-2000" ...
## .. attr(*, "index")=Classes 'pindex' and 'data.frame': 4596 obs. of 2 variables:
## .. ..$ id : Factor w/ 1149 levels "1","2","3","4",...: 1 1 1 1 2 2 2 2 3 3 ...
## .. ..$ year: Factor w/ 4 levels "1997","1998",...: 1 2 3 4 1 2 3 4 1 2 ...
## $ id : Factor w/ 1149 levels "1","2","3","4",...: 1 1 1 1 2 2 2 2 3 3 ...
## .. attr(*, "names")= chr "1-1997" "1-1998" "1-1999" "1-2000" ...
## .. attr(*, "index")=Classes 'pindex' and 'data.frame': 4596 obs. of 2 variables:
## .. ..$ id : Factor w/ 1149 levels "1","2","3","4",...: 1 1 1 1 2 2 2 2 3 3 ...
## .. ..$ year: Factor w/ 4 levels "1997","1998",...: 1 2 3 4 1 2 3 4 1 2 ...
## $ dist : 'pseries' Named int 528 528 528 528 861 861 861 861 852 852 ...
## .. attr(*, "names")= chr "1-1997" "1-1998" "1-1999" "1-2000" ...
## .. attr(*, "index")=Classes 'pindex' and 'data.frame': 4596 obs. of 2 variables:
## .. ..$ id : Factor w/ 1149 levels "1","2","3","4",...: 1 1 1 1 2 2 2 2 3 3 ...
## .. ..$ year: Factor w/ 4 levels "1997","1998",...: 1 2 3 4 1 2 3 4 1 2 ...
## $ passen : 'pseries' Named int 152 265 336 298 282 178 204 190 241 253 ...
## .. attr(*, "names")= chr "1-1997" "1-1998" "1-1999" "1-2000" ...
## .. attr(*, "index")=Classes 'pindex' and 'data.frame': 4596 obs. of 2 variables:
## .. ..$ id : Factor w/ 1149 levels "1","2","3","4",...: 1 1 1 1 2 2 2 2 3 3 ...
## .. ..$ year: Factor w/ 4 levels "1997","1998",...: 1 2 3 4 1 2 3 4 1 2 ...
## $ fare : 'pseries' Named int 106 106 113 123 104 105 115 129 207 188 ...
## .. attr(*, "names")= chr "1-1997" "1-1998" "1-1999" "1-2000" ...
## .. attr(*, "index")=Classes 'pindex' and 'data.frame': 4596 obs. of 2 variables:
## .. ..$ id : Factor w/ 1149 levels "1","2","3","4",...: 1 1 1 1 2 2 2 2 3 3 ...
## .. ..$ year: Factor w/ 4 levels "1997","1998",...: 1 2 3 4 1 2 3 4 1 2 ...
## $ bmktshr: 'pseries' Named num 0.839 0.813 0.826 0.861 0.58 ...
## .. attr(*, "names")= chr "1-1997" "1-1998" "1-1999" "1-2000" ...
## .. attr(*, "index")=Classes 'pindex' and 'data.frame': 4596 obs. of 2 variables:
## .. ..$ id : Factor w/ 1149 levels "1","2","3","4",...: 1 1 1 1 2 2 2 2 3 3 ...
## .. ..$ year: Factor w/ 4 levels "1997","1998",...: 1 2 3 4 1 2 3 4 1 2 ...
## $ ldist : 'pseries' Named num 6.27 6.27 6.27 6.27 6.76 ...
## .. attr(*, "names")= chr "1-1997" "1-1998" "1-1999" "1-2000" ...
## .. attr(*, "index")=Classes 'pindex' and 'data.frame': 4596 obs. of 2 variables:
## .. ..$ id : Factor w/ 1149 levels "1","2","3","4",...: 1 1 1 1 2 2 2 2 3 3 ...
## .. ..$ year: Factor w/ 4 levels "1997","1998",...: 1 2 3 4 1 2 3 4 1 2 ...
## $ y98 : 'pseries' Named int 0 1 0 0 0 1 0 0 0 1 ...
## .. attr(*, "names")= chr "1-1997" "1-1998" "1-1999" "1-2000" ...
## .. attr(*, "index")=Classes 'pindex' and 'data.frame': 4596 obs. of 2 variables:
## .. ..$ id : Factor w/ 1149 levels "1","2","3","4",...: 1 1 1 1 2 2 2 2 3 3 ...
## .. ..$ year: Factor w/ 4 levels "1997","1998",...: 1 2 3 4 1 2 3 4 1 2 ...
## $ y99 : 'pseries' Named int 0 0 1 0 0 0 1 0 0 0 ...
## .. attr(*, "names")= chr "1-1997" "1-1998" "1-1999" "1-2000" ...
## .. attr(*, "index")=Classes 'pindex' and 'data.frame': 4596 obs. of 2 variables:
## .. ..$ id : Factor w/ 1149 levels "1","2","3","4",...: 1 1 1 1 2 2 2 2 3 3 ...
## .. ..$ year: Factor w/ 4 levels "1997","1998",...: 1 2 3 4 1 2 3 4 1 2 ...
## $ y00 : 'pseries' Named int 0 0 0 1 0 0 0 1 0 0 ...
## .. attr(*, "names")= chr "1-1997" "1-1998" "1-1999" "1-2000" ...
## .. attr(*, "index")=Classes 'pindex' and 'data.frame': 4596 obs. of 2 variables:
## .. ..$ id : Factor w/ 1149 levels "1","2","3","4",...: 1 1 1 1 2 2 2 2 3 3 ...
## .. ..$ year: Factor w/ 4 levels "1997","1998",...: 1 2 3 4 1 2 3 4 1 2 ...
## $ lfare : 'pseries' Named num 4.66 4.66 4.73 4.81 4.64 ...
## .. attr(*, "names")= chr "1-1997" "1-1998" "1-1999" "1-2000" ...
## .. attr(*, "index")=Classes 'pindex' and 'data.frame': 4596 obs. of 2 variables:
## .. ..$ id : Factor w/ 1149 levels "1","2","3","4",...: 1 1 1 1 2 2 2 2 3 3 ...
## .. ..$ year: Factor w/ 4 levels "1997","1998",...: 1 2 3 4 1 2 3 4 1 2 ...
## $ ldistsq: 'pseries' Named num 39.3 39.3 39.3 39.3 45.7 ...
## .. attr(*, "names")= chr "1-1997" "1-1998" "1-1999" "1-2000" ...
## .. attr(*, "index")=Classes 'pindex' and 'data.frame': 4596 obs. of 2 variables:
## .. ..$ id : Factor w/ 1149 levels "1","2","3","4",...: 1 1 1 1 2 2 2 2 3 3 ...
## .. ..$ year: Factor w/ 4 levels "1997","1998",...: 1 2 3 4 1 2 3 4 1 2 ...

```



```
## $ concen : 'pseries' Named num 0.839 0.813 0.826 0.861 0.58 ...
## ..- attr(*, "names")= chr "1-1997" "1-1998" "1-1999" "1-2000" ...
## ..- attr(*, "index")=Classes 'pindex' and 'data.frame': 4596 obs. of 2 variables:
## .. ..$ id : Factor w/ 1149 levels "1","2","3","4",...: 1 1 1 1 2 2 2 2 3 3 ...
## .. ..$ year: Factor w/ 4 levels "1997","1998",...: 1 2 3 4 1 2 3 4 1 2 ...
## $ lpassen: 'pseries' Named num 5.02 5.58 5.82 5.7 5.64 ...
## ..- attr(*, "names")= chr "1-1997" "1-1998" "1-1999" "1-2000" ...
## ..- attr(*, "index")=Classes 'pindex' and 'data.frame': 4596 obs. of 2 variables:
## .. ..$ id : Factor w/ 1149 levels "1","2","3","4",...: 1 1 1 1 2 2 2 2 3 3 ...
## .. ..$ year: Factor w/ 4 levels "1997","1998",...: 1 2 3 4 1 2 3 4 1 2 ...
## - attr(*, "time.stamp")= chr "25 Jun 2011 23:03"
## - attr(*, "index")=Classes 'pindex' and 'data.frame': 4596 obs. of 2 variables:
## ..$ id : Factor w/ 1149 levels "1","2","3","4",...: 1 1 1 1 2 2 2 2 3 3 ...
## ..$ year: Factor w/ 4 levels "1997","1998",...: 1 2 3 4 1 2 3 4 1 2 ...
```

```
head(pdata,n = 20)
```

```
##      year id dist passen fare bmktsshr   ldist y98 y99 y00   lfare
## 1-1997 1997 1 528 152 106 0.8386 6.269096 0 0 0 4.663439
## 1-1998 1998 1 528 265 106 0.8133 6.269096 1 0 0 4.663439
## 1-1999 1999 1 528 336 113 0.8262 6.269096 0 1 0 4.727388
## 1-2000 2000 1 528 298 123 0.8612 6.269096 0 0 1 4.812184
## 2-1997 1997 2 861 282 104 0.5798 6.758094 0 0 0 4.644391
## 2-1998 1998 2 861 178 105 0.5817 6.758094 1 0 0 4.653960
## 2-1999 1999 2 861 204 115 0.7319 6.758094 0 1 0 4.744932
## 2-2000 2000 2 861 190 129 0.5386 6.758094 0 0 1 4.859812
## 3-1997 1997 3 852 241 207 0.8180 6.747587 0 0 0 5.332719
## 3-1998 1998 3 852 253 188 0.8172 6.747587 1 0 0 5.236442
## 3-1999 1999 3 852 244 229 0.7998 6.747587 0 1 0 5.433722
## 3-2000 2000 3 852 244 247 0.7097 6.747587 0 0 1 5.509388
## 4-1997 1997 4 724 233 243 0.4604 6.584791 0 0 0 5.493062
## 4-1998 1998 4 724 248 226 0.4614 6.584791 1 0 0 5.420535
## 4-1999 1999 4 724 224 229 0.4334 6.584791 0 1 0 5.433722
## 4-2000 2000 4 724 339 176 0.3716 6.584791 0 0 1 5.170484
## 5-1997 1997 5 1073 396 119 0.4571 6.978214 0 0 0 4.779123
## 5-1998 1998 5 1073 529 105 0.5632 6.978214 1 0 0 4.653960
## 5-1999 1999 5 1073 494 121 0.5008 6.978214 0 1 0 4.795791
## 5-2000 2000 5 1073 710 109 0.5197 6.978214 0 0 1 4.691348
##      ldistsq concen lpassen
## 1-1997 39.30157 0.8386 5.023880
## 1-1998 39.30157 0.8133 5.579730
## 1-1999 39.30157 0.8262 5.817111
## 1-2000 39.30157 0.8612 5.697093
## 2-1997 45.67184 0.5798 5.641907
## 2-1998 45.67184 0.5817 5.181784
## 2-1999 45.67184 0.7319 5.318120
## 2-2000 45.67184 0.5386 5.247024
## 3-1997 45.52993 0.8180 5.484797
## 3-1998 45.52993 0.8172 5.533390
## 3-1999 45.52993 0.7998 5.497168
## 3-2000 45.52993 0.7097 5.497168
## 4-1997 43.35947 0.4604 5.451038
## 4-1998 43.35947 0.4614 5.513429
## 4-1999 43.35947 0.4334 5.411646
## 4-2000 43.35947 0.3716 5.826000
## 5-1997 48.69547 0.4571 5.981414
## 5-1998 48.69547 0.5632 6.270988
## 5-1999 48.69547 0.5008 6.202536
## 5-2000 48.69547 0.5197 6.565265
```

```
# existen diferentes alternativas para definir un panel
# Esta quita los índices del dataframe (dejan de estar como columnas)
pdata <- pdata.frame(airfare, index=c("id","year"),drop.index=TRUE, row.names = TRUE)
head(pdata,n = 20)
```

```
##      dist passen fare bmktshr   ldist y98 y99 y00   lfare ldistsq
## 1-1997  528   152  106  0.8386 6.269096  0  0  0  4.663439 39.30157
## 1-1998  528   265  106  0.8133 6.269096  1  0  0  4.663439 39.30157
## 1-1999  528   336  113  0.8262 6.269096  0  1  0  4.727388 39.30157
## 1-2000  528   298  123  0.8612 6.269096  0  0  1  4.812184 39.30157
## 2-1997  861   282  104  0.5798 6.758094  0  0  0  4.644391 45.67184
## 2-1998  861   178  105  0.5817 6.758094  1  0  0  4.653960 45.67184
## 2-1999  861   204  115  0.7319 6.758094  0  1  0  4.744932 45.67184
## 2-2000  861   190  129  0.5386 6.758094  0  0  1  4.859812 45.67184
## 3-1997  852   241  207  0.8180 6.747587  0  0  0  5.332719 45.52993
## 3-1998  852   253  188  0.8172 6.747587  1  0  0  5.236442 45.52993
## 3-1999  852   244  229  0.7998 6.747587  0  1  0  5.433722 45.52993
## 3-2000  852   244  247  0.7097 6.747587  0  0  1  5.509388 45.52993
## 4-1997  724   233  243  0.4604 6.584791  0  0  0  5.493062 43.35947
## 4-1998  724   248  226  0.4614 6.584791  1  0  0  5.420535 43.35947
## 4-1999  724   224  229  0.4334 6.584791  0  1  0  5.433722 43.35947
## 4-2000  724   339  176  0.3716 6.584791  0  0  1  5.170484 43.35947
## 5-1997 1073   396  119  0.4571 6.978214  0  0  0  4.779123 48.69547
## 5-1998 1073   529  105  0.5632 6.978214  1  0  0  4.653960 48.69547
## 5-1999 1073   494  121  0.5008 6.978214  0  1  0  4.795791 48.69547
## 5-2000 1073   710  109  0.5197 6.978214  0  0  1  4.691348 48.69547
##      concen  lpassen
## 1-1997 0.8386 5.023880
## 1-1998 0.8133 5.579730
## 1-1999 0.8262 5.817111
## 1-2000 0.8612 5.697093
## 2-1997 0.5798 5.641907
## 2-1998 0.5817 5.181784
## 2-1999 0.7319 5.318120
## 2-2000 0.5386 5.247024
## 3-1997 0.8180 5.484797
## 3-1998 0.8172 5.533390
## 3-1999 0.7998 5.497168
## 3-2000 0.7097 5.497168
## 4-1997 0.4604 5.451038
## 4-1998 0.4614 5.513429
## 4-1999 0.4334 5.411646
## 4-2000 0.3716 5.826000
## 5-1997 0.4571 5.981414
## 5-1998 0.5632 6.270988
## 5-1999 0.5008 6.202536
## 5-2000 0.5197 6.565265
```

```
#str(pdata)
# Esta NO quita los índices del dataframe (continúan estando como columnas)
pdata <- pdata.frame(airfare, index=c("id","year"),drop.index=FALSE, row.names = TRUE)
head(pdata,n = 20)
```

```
##      year id dist passen fare bmktsshr      ldist y98 y99 y00      lfare
## 1-1997 1997  1  528    152   106   0.8386 6.269096    0  0  0 4.663439
## 1-1998 1998  1  528    265   106   0.8133 6.269096    1  0  0 4.663439
## 1-1999 1999  1  528    336   113   0.8262 6.269096    0  1  0 4.727388
## 1-2000 2000  1  528    298   123   0.8612 6.269096    0  0  1 4.812184
## 2-1997 1997  2  861    282   104   0.5798 6.758094    0  0  0 4.644391
## 2-1998 1998  2  861    178   105   0.5817 6.758094    1  0  0 4.653960
## 2-1999 1999  2  861    204   115   0.7319 6.758094    0  1  0 4.744932
## 2-2000 2000  2  861    190   129   0.5386 6.758094    0  0  1 4.859812
## 3-1997 1997  3  852    241   207   0.8180 6.747587    0  0  0 5.332719
## 3-1998 1998  3  852    253   188   0.8172 6.747587    1  0  0 5.236442
## 3-1999 1999  3  852    244   229   0.7998 6.747587    0  1  0 5.433722
## 3-2000 2000  3  852    244   247   0.7097 6.747587    0  0  1 5.509388
## 4-1997 1997  4  724    233   243   0.4604 6.584791    0  0  0 5.493062
## 4-1998 1998  4  724    248   226   0.4614 6.584791    1  0  0 5.420535
## 4-1999 1999  4  724    224   229   0.4334 6.584791    0  1  0 5.433722
## 4-2000 2000  4  724    339   176   0.3716 6.584791    0  0  1 5.170484
## 5-1997 1997  5 1073    396   119   0.4571 6.978214    0  0  0 4.779123
## 5-1998 1998  5 1073    529   105   0.5632 6.978214    1  0  0 4.653960
## 5-1999 1999  5 1073    494   121   0.5008 6.978214    0  1  0 4.795791
## 5-2000 2000  5 1073    710   109   0.5197 6.978214    0  0  1 4.691348
##      ldistsq concen  lpassen
## 1-1997 39.30157 0.8386 5.023880
## 1-1998 39.30157 0.8133 5.579730
## 1-1999 39.30157 0.8262 5.817111
## 1-2000 39.30157 0.8612 5.697093
## 2-1997 45.67184 0.5798 5.641907
## 2-1998 45.67184 0.5817 5.181784
## 2-1999 45.67184 0.7319 5.318120
## 2-2000 45.67184 0.5386 5.247024
## 3-1997 45.52993 0.8180 5.484797
## 3-1998 45.52993 0.8172 5.533390
## 3-1999 45.52993 0.7998 5.497168
## 3-2000 45.52993 0.7097 5.497168
## 4-1997 43.35947 0.4604 5.451038
## 4-1998 43.35947 0.4614 5.513429
## 4-1999 43.35947 0.4334 5.411646
## 4-2000 43.35947 0.3716 5.826000
## 5-1997 48.69547 0.4571 5.981414
## 5-1998 48.69547 0.5632 6.270988
## 5-1999 48.69547 0.5008 6.202536
## 5-2000 48.69547 0.5197 6.565265
```

```
#str(pdata)
```

```
# para extraer los índices
head(attr(pdata, "index"),n = 20)
```

```
##      id year
## 1     1 1997
## 2     1 1998
## 3     1 1999
## 4     1 2000
## 5     2 1997
## 6     2 1998
## 7     2 1999
## 8     2 2000
## 9     3 1997
## 10    3 1998
## 11    3 1999
## 12    3 2000
## 13    4 1997
## 14    4 1998
## 15    4 1999
## 16    4 2000
## 17    5 1997
## 18    5 1998
## 19    5 1999
## 20    5 2000
```

## 4 Estadística descriptiva del panel de datos

```
# Descriptive statistics
# Para obtener información sobre porcentaje de la variabilidad que proviene de id y de year
summary(pdata$fare)
```

```
## total sum of squares: 25765270
##           id           time
## 0.944152346 0.005463992
```

```
head(pdata$fare,n=20)
```

```
## 1-1997 1-1998 1-1999 1-2000 2-1997 2-1998 2-1999 2-2000 3-1997 3-1998
##    106    106    113    123    104    105    115    129    207    188
## 3-1999 3-2000 4-1997 4-1998 4-1999 4-2000 5-1997 5-1998 5-1999 5-2000
##    229    247    243    226    229    176    119    105    121    109
```

```
#Otra forma de mostrar los datos
head(as.matrix(pdata$fare))
```

```
##      1997 1998 1999 2000
## 1    106  106  113  123
## 2    104  105  115  129
## 3    207  188  229  247
## 4    243  226  229  176
## 5    119  105  121  109
## 6    181  197  190  171
```

```
# between (con minúscula) proporciona la media temporal de cada individuo
# y proporciona por tanto n datos (n= número de individuos)
```

```
head(between(pdata$fare),n=20) # media de cada individuo
```

```
##      1      2      3      4      5      6      7      8      9     10
## 112.00 113.25 217.75 218.50 113.50 184.75 216.75 148.25 154.50 237.00
##      11     12     13     14     15     16     17     18     19     20
## 183.50 108.25 176.00  71.50 142.00 139.25 102.25 116.50 192.75 244.75
```

```
# Between (con mayúscula) proporciona también la media temporal de cada individuo
# pero ahora la salida es de tamaño n*T
head(Between(pdata$fare),n=20) # media de cada individuo
```

```
##      1      1      1      1      2      2      2      2      3      3
## 112.00 112.00 112.00 112.00 113.25 113.25 113.25 113.25 217.75 217.75
##      3      3      4      4      4      4      5      5      5      5
## 217.75 217.75 218.50 218.50 218.50 218.50 113.50 113.50 113.50 113.50
```

```
#Within (con mayuscula) transforma cada serie restándole a cada individuo su media temporal
head(Within(pdata$fare),n=20) # dato del individuo en diferencias respecto a la media de cada individuo
```

```
## 1-1997 1-1998 1-1999 1-2000 2-1997 2-1998 2-1999 2-2000 3-1997 3-1998
## -6.00 -6.00  1.00 11.00 -9.25 -8.25  1.75 15.75 -10.75 -29.75
## 3-1999 3-2000 4-1997 4-1998 4-1999 4-2000 5-1997 5-1998 5-1999 5-2000
## 11.25 29.25 24.50  7.50 10.50 -42.50  5.50 -8.50  7.50 -4.50
```

```
# diff(), lag() y lead() calcula la primera diferencia temporal,
# lag el retardo (para atr?s con k>0), y lead(el adelanto temporal con k>0)
# pero teniendo en cuenta la estructura de datos de panel
```

```
head(diff(pdata$fare,lag=1),n=20) # media de cada individuo
```

```
## 1-1997 1-1998 1-1999 1-2000 2-1997 2-1998 2-1999 2-2000 3-1997 3-1998
##      NA      0      7      10      NA      1      10      14      NA     -19
## 3-1999 3-2000 4-1997 4-1998 4-1999 4-2000 5-1997 5-1998 5-1999 5-2000
##      41      18      NA     -17      3     -53      NA     -14      16     -12
```

```
head(lag(pdata$fare,k=1),n=20) # media de cada individuo
```

```
## 1-1997 1-1998 1-1999 1-2000 2-1997 2-1998 2-1999 2-2000 3-1997 3-1998
##      NA     106     106     113      NA     104     105     115      NA     207
## 3-1999 3-2000 4-1997 4-1998 4-1999 4-2000 5-1997 5-1998 5-1999 5-2000
##      188      229      NA     243     226     229      NA     119     105     121
```

```
head(lead(pdata$fare,k=1),n=20) # media de cada individuo
```

##	1-1997	1-1998	1-1999	1-2000	2-1997	2-1998	2-1999	2-2000	3-1997	3-1998
##	106	113	123	NA	105	115	129	NA	188	229
##	3-1999	3-2000	4-1997	4-1998	4-1999	4-2000	5-1997	5-1998	5-1999	5-2000
##	247	NA	226	229	176	NA	105	121	109	NA

## 5 Métodos de estimación del modelo lineal con datos de panel

Estimación de los modelos con datos de Panel.

Existen 4 métodos en la librería `plm`

1. **pvcmm**: para modelos con todos los coeficientes variables;
2. **plm**: para modelos “between” efectos fijos o “within”, “pooling” todos los coeficientes iguales, “fd” first diferencias, “random” efectos aleatorios (GLS);
3. **pgmm**: método de los momentos por ejemplo para modelos dinámicos Arellano-Bond (no los veremos)
4. **pggls**: para ganar precisión estimando varianzas por cluster de individuos

Por otra parte se pueden incluir “effect” **efectos individuales** (fijos en el tiempo), **efectos temporales** (fijos entre individuos), o ambos

```
# Estimación de modelos con datos de panel: fórmula general

# plm(formula, data, subset, weights, na.action,
#   effect = c("individual", "time", "twoways", "nested"),
#   model = c("within", "random", "ht", "between", "pooling", "fd"))
```

### 5.1 Pooled OLS estimator

No considera datos de panel, es como poner todos los datos seguidos y considerar una constante y todas las pendientes (las betas) iguales

Consideremos inicialmente un modelo donde sólo aparecerá el índice de concentración de mercado `concen`, la distancia que recorrerá el vuelo `ldist`, su cuadrado, y efectos de tiempo (efectos anuales iguales para todos los individuos).

```
pooling <- plm(lfare ~ concen + ldist + I(ldist^2) + y98 + y99 + y00, data=pdata, model= "pooling")
summary(pooling)
```

```
## Pooling Model
##
## Call:
## plm(formula = lfare ~ concen + ldist + I(ldist^2) + y98 + y99 +
##      y00, data = pdata, model = "pooling")
##
## Balanced Panel: n = 1149, T = 4, N = 4596
##
## Residuals:
##      Min.      1st Qu.      Median      3rd Qu.      Max.
## -1.307955 -0.253740 -0.026729  0.247446  0.958113
##
## Coefficients:
##              Estimate Std. Error t-value Pr(>|t|)
## (Intercept)  6.2092570   0.4206247  14.7620 < 2.2e-16 ***
## concen       0.3601203   0.0300691  11.9764 < 2.2e-16 ***
## ldist       -0.9016002   0.1282730  -7.0288 2.391e-12 ***
## I(ldist^2)   0.1030196   0.0097255  10.5927 < 2.2e-16 ***
## y98          0.0211244   0.0140419   1.5044  0.132553
## y99          0.0378496   0.0140413   2.6956  0.007052 **
## y00          0.0998700   0.0140432   7.1116 1.324e-12 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Total Sum of Squares:    875.09
## Residual Sum of Squares: 519.64
## R-Squared:              0.40619
## Adj. R-Squared: 0.40541
## F-statistic: 523.175 on 6 and 4589 DF, p-value: < 2.22e-16
```

```
stargazer(pooling,type = "text")
```

```
##
## =====
##           Dependent variable:
##           -----
##           lfare
## -----
## concen           0.360***
##                  (0.030)
##
## ldist            -0.902***
##                  (0.128)
##
## I(ldist2)         0.103***
##                  (0.010)
##
## y98               0.021
##                  (0.014)
##
## y99               0.038***
##                  (0.014)
##
## y00               0.100***
##                  (0.014)
##
## Constant          6.209***
##                  (0.421)
##
## -----
## Observations      4,596
## R2                 0.406
## Adjusted R2        0.405
## F Statistic    523.175*** (df = 6; 4589)
## =====
## Note:             *p<0.1; **p<0.05; ***p<0.01
```

Darí­a el mismo resultado que estimar por lm directamente del data.frame inicial de datos

```
pooling2 <- lm(lfare ~ concen + ldist + I(ldist^2) + y98 + y99 + y00, data=airfare)
stargazer(pooling,pooling2,type = "text")
```



```
##
## =====
##                               Dependent variable:
##                               -----
##                               lfare
##                               panel      OLS
##                               linear
##                               (1)        (2)
## -----
## concen                0.360***      0.360***
##                        (0.030)      (0.030)
##
## ldist                 -0.902***      -0.902***
##                        (0.128)      (0.128)
##
## I(ldist2)              0.103***      0.103***
##                        (0.010)      (0.010)
##
## y98                    0.021         0.021
##                        (0.014)      (0.014)
##
## y99                    0.038***      0.038***
##                        (0.014)      (0.014)
##
## y00                    0.100***      0.100***
##                        (0.014)      (0.014)
##
## Constant              6.209***      6.209***
##                        (0.421)      (0.421)
##
## -----
## Observations          4,596         4,596
## R2                    0.406         0.406
## Adjusted R2           0.405         0.405
## Residual Std. Error    0.337 (df = 4589)
## F Statistic (df = 6; 4589) 523.175*** 523.175***
## =====
## Note:                  *p<0.1; **p<0.05; ***p<0.01
```

## 6 Between estimator

Between Estima el modelo pero con las medias de cada id, es una estimación con datos de sección cruzada

El problema de este estimador es que estima  $y_{bar} = \beta x_{bar} + \beta_0 + \epsilon_i + u_i$ , pero el nuevo término de error  $\epsilon_i + u_i$  es no observable y es el error compuesto, el problema del estimador Between es que estará sesgado si  $Cov(x, \epsilon_i)$  no es cero y si es cero, tampoco es buen estimador, mejor utilizar Random effects porque Between no recoge bien el efecto de las variables que varían en el tiempo

```
between <- plm(lfare ~ concen + ldist + I(ldist^2) + y98 + y99 + y00, data=pdata, model= "between")
summary(between)
```

```
## Oneway (individual) effect Between Model
##
## Call:
## plm(formula = lfare ~ concen + ldist + I(ldist^2) + y98 + y99 +
##      y00, data = pdata, model = "between")
##
## Balanced Panel: n = 1149, T = 4, N = 4596
## Observations used in estimation: 1149
##
## Residuals:
##      Min.      1st Qu.      Median      3rd Qu.      Max.
## -0.723069 -0.243467 -0.023001  0.225211  0.904181
##
## Coefficients:
##              Estimate Std. Error t-value Pr(>|t|)
## (Intercept)  6.247135   0.809189   7.7202 2.520e-14 ***
## concen       0.382494   0.061149   6.2551 5.599e-10 ***
## ldist      -0.908930   0.246905  -3.6813 0.0002429 ***
## I(ldist^2)   0.103843   0.018728   5.5448 3.651e-08 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Total Sum of Squares:    207.48
## Residual Sum of Squares: 120.01
## R-Squared:      0.42158
## Adj. R-Squared: 0.42007
## F-statistic: 278.178 on 3 and 1145 DF, p-value: < 2.22e-16
```

```
stargazer(pooling, between,type = "text")
```

```

##
## =====
##                               Dependent variable:
##                               -----
##                               lfare
##                               (1)                (2)
## -----
## concen                0.360***                0.382***
##                        (0.030)                (0.061)
##
## ldist                 -0.902***                -0.909***
##                        (0.128)                (0.247)
##
## I(ldist2)             0.103***                0.104***
##                        (0.010)                (0.019)
##
## y98                   0.021
##                        (0.014)
##
## y99                   0.038***
##                        (0.014)
##
## y00                   0.100***
##                        (0.014)
##
## Constant              6.209***                6.247***
##                        (0.421)                (0.809)
##
## -----
## Observations          4,596                1,149
## R2                    0.406                0.422
## Adjusted R2           0.405                0.420
## F Statistic  523.175*** (df = 6; 4589) 278.178*** (df = 3; 1145)
## =====
## Note:                  *p<0.1; **p<0.05; ***p<0.01

```

## 7 Estimador cuando hay un problema de endogeneidad con los efectos fijos no observables

Tenemos el modelo:

$$y_{it} = \beta x_{it} + \beta_0 + c_i + u_{it},$$

$c_i$  es no observable y puede considerarse como parte del error compuesto, por tanto la estimación de  $\beta$  dependerá de si  $\text{Cov}(x_{it}, c_i) \neq 0$  para alguna de las variables explicativas. Para estos casos  $\text{Cov}(x_{it}, c_i) \neq 0$  tenemos dos estimadores consistentes. Estos dos estimadores lo que hacen es transformar el modelo para que desaparezcan los efectos fijos  $c_i$  que causan el problema el problema de inconsistencia (en lugar, por ejemplo, de buscar instrumentos para  $x_{it}$  y estimar por variables instrumentales). El problema en ambas es que todas aquellas variables fijas en el tiempo desaparecen, no solo los efectos heterogeneos no observables, también las observables (como la raza, el sexo, nacionalidad, etc). Más adelante proponemos dos alternativas para intentar recoger estos efectos fijos sí observables

### 7.1 Estimador de Primeras diferencias y de Efectos fijos

```
# First differences estimator
firstdiff <- plm(lfare ~ concen + ldist + I(ldist^2) + y98 + y99 + y00, data=pdata, model= "fd")
summary(firstdiff)
```

```
## Oneway (individual) effect First-Difference Model
##
## Call:
## plm(formula = lfare ~ concen + ldist + I(ldist^2) + y98 + y99 +
##      y00, data = pdata, model = "fd")
##
## Balanced Panel: n = 1149, T = 4, N = 4596
## Observations used in estimation: 3447
##
## Residuals:
##      Min.      1st Qu.      Median      3rd Qu.      Max.
## -0.8035456 -0.0554571  0.0021427  0.0583380  1.0958535
##
## Coefficients: (1 dropped because of singularities)
##              Estimate Std. Error t-value Pr(>|t|)
## (Intercept)  0.0326166  0.0021330 15.2915 < 2.2e-16 ***
## concen      0.1759764  0.0284387  6.1879 6.813e-10 ***
## y98         -0.0098474  0.0030341 -3.2456 0.001183 **
## y99         -0.0287967  0.0030129 -9.5577 < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Total Sum of Squares:    56.008
## Residual Sum of Squares: 53.867
## R-Squared:              0.038223
## Adj. R-Squared: 0.037385
## F-statistic: 45.6104 on 3 and 3443 DF, p-value: < 2.22e-16
```

## 7.2 Estimador de efectos Fijos

```
# Fixed effects or within estimator
fixed <- plm(lfare ~ concen + ldist + I(ldist^2) + y98 + y99 + y00, data=pdata, model= "within")
summary(fixed)
```

```
## Oneway (individual) effect Within Model
##
## Call:
## plm(formula = lfare ~ concen + ldist + I(ldist^2) + y98 + y99 +
##      y00, data = pdata, model = "within")
##
## Balanced Panel: n = 1149, T = 4, N = 4596
##
## Residuals:
##      Min.      1st Qu.      Median      3rd Qu.      Max.
## -0.8812810 -0.0397341  0.0023907  0.0436350  0.9188102
##
## Coefficients:
##      Estimate Std. Error t-value Pr(>|t|)
## concen 0.1688590  0.0294101  5.7415 1.020e-08 ***
## y98     0.0228328  0.0044515  5.1292 3.071e-07 ***
## y99     0.0363819  0.0044495  8.1766 4.061e-16 ***
## y00     0.0977717  0.0044555 21.9441 < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Total Sum of Squares:    45.169
## Residual Sum of Squares: 39.06
## R-Squared:    0.13524
## Adj. R-Squared: -0.1541
## F-statistic: 134.611 on 4 and 3443 DF, p-value: < 2.22e-16
```

## 8 Comparación de modelos

```
stargazer(pooling, between, fixed ,type = "text")
```

```
##
## =====
##                               Dependent variable:
##                               -----
##                               lfare
##                               (1)          (2)          (3)
## -----
## concen          0.360***          0.382***          0.169***
##                  (0.030)          (0.061)          (0.029)
##
## ldist           -0.902***          -0.909***
##                  (0.128)          (0.247)
##
## I(ldist2)        0.103***          0.104***
##                  (0.010)          (0.019)
##
## y98              0.021              0.023***
##                  (0.014)              (0.004)
##
## y99              0.038***              0.036***
##                  (0.014)              (0.004)
##
## y00              0.100***              0.098***
##                  (0.014)              (0.004)
##
## Constant         6.209***          6.247***
##                  (0.421)          (0.809)
##
## -----
## Observations      4,596              1,149              4,596
## R2                 0.406              0.422              0.135
## Adjusted R2        0.405              0.420              -0.154
## F Statistic  523.175*** (df = 6; 4589) 278.178*** (df = 3; 1145) 134.611*** (df = 4; 3443)
## =====
## Note:                                     *p<0.1; **p<0.05; ***p<0.01
```

## 9 Efectos temporales

Los efectos temporales se pueden estimar sin variables dummies introduciendo la opción de efectos temporales en la estimación del modelo:

```
fixed.2 <- plm(lfare ~ concen + ldist + I(ldist^2), data=pdata, model= "within", effect = "twoways")
summary(fixed.2)
```

```
## Twoways effects Within Model
##
## Call:
## plm(formula = lfare ~ concen + ldist + I(ldist^2), data = pdata,
##     effect = "twoways", model = "within")
##
## Balanced Panel: n = 1149, T = 4, N = 4596
##
## Residuals:
##      Min.      1st Qu.      Median      3rd Qu.      Max.
## -0.8812810 -0.0397341  0.0023907  0.0436350  0.9188102
##
## Coefficients:
##      Estimate Std. Error t-value Pr(>|t|)
## concen   0.16886    0.02941  5.7415 1.02e-08 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Total Sum of Squares:    39.434
## Residual Sum of Squares: 39.06
## R-Squared:      0.0094837
## Adj. R-Squared: -0.32194
## F-statistic: 32.9651 on 1 and 3443 DF, p-value: 1.0196e-08
```

```
# para extraer los efectos fijos
head(summary(fixef(fixed, type="level")),10)
```

```
##      Estimate Std. Error  t-value Pr(>|t|)
## 1  4.536398 0.05873607 77.23360      0
## 2  4.583861 0.05626634 81.46720      0
## 3  5.206069 0.05815075 89.52710      0
## 4  5.267308 0.05483261 96.06159      0
## 5  4.604657 0.05541569 83.09302      0
## 6  5.021535 0.05994560 83.76821      0
## 7  5.225488 0.05685448 91.90987      0
## 8  4.854573 0.05627291 86.26838      0
## 9  4.961961 0.05375440 92.30800      0
## 10 5.378720 0.05393020 99.73483      0
```

```
head(summary(fixef(fixed, type="dmean")),10) #dmean desviaciones respecto a la media global
```

```
##      Estimate Std. Error  t-value      Pr(>|t|)
## 1  -0.416932156 0.05873607 -7.0984003 1.525839e-12
## 2  -0.369469489 0.05626634 -6.5664392 5.927915e-11
## 3   0.252737986 0.05815075  4.3462547 1.424854e-05
## 4   0.313977095 0.05483261  5.7261019 1.115743e-08
## 5  -0.348673441 0.05541569 -6.2919626 3.529473e-10
## 6   0.068204739 0.05994560  1.1377773 2.552926e-01
## 7   0.272157189 0.05685448  4.7869085 1.764975e-06
## 8  -0.098757654 0.05627291 -1.7549768 7.935220e-02
## 9   0.008630838 0.05375440  0.1605606 8.724489e-01
## 10  0.425389167 0.05393020  7.8877723 4.097373e-15
```

```
# cuando tengo los dos efectos
fixef(fixed.2, effect = "time", type="level")
```

```
##      1997      1998      1999      2000
## 4.953331 4.976163 4.989712 5.051102
```

```
summary(fixef(fixed.2, effect = "time", type="dmean"))
```

```
##      Estimate Std. Error t-value Pr(>|t|)
## 1997 -0.0392466  0.0182869 -2.1462 0.031931 *
## 1998 -0.0164138  0.0185458 -0.8850 0.376196
## 1999 -0.0028647  0.0180647 -0.1586 0.874008
## 2000  0.0585251  0.0179692  3.2570 0.001137 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

## 10 Estimador de Efectos Aleatorios

Para los casos en los que **NO** tengo un problema de endogeneidad  $Cov(x_{it}, \epsilon_{it})=0$  entonces no hay problemas de inconsistencia al utilizar MCO, aunque sí de eficiencia, porque la matriz de Var-Cov del término de error no será la matriz identidad. Utilizando el estimador de MCG obtendríamos un estimador no sólo insesgado sino también eficiente

```
# Random effects estimator
random <- plm(lfare ~ concen + ldist + I(ldist^2) + y98 + y99 + y00, data=pdata, model= "random")
summary(random)
```



```

## Oneway (individual) effect Random Effect Model
##   (Swamy-Arora's transformation)
##
## Call:
## plm(formula = lfare ~ concen + ldist + I(ldist^2) + y98 + y99 +
##       y00, data = pdata, model = "random")
##
## Balanced Panel: n = 1149, T = 4, N = 4596
##
## Effects:
##               var std.dev share
## idiosyncratic 0.01134 0.10651  0.1
## individual    0.10198 0.31934  0.9
## theta: 0.8355
##
## Residuals:
##      Min.      1st Qu.      Median      3rd Qu.      Max.
## -0.8316339 -0.0639906 -0.0037195  0.0626552  0.8655403
##
## Coefficients:
##              Estimate Std. Error z-value Pr(>|z|)
## (Intercept)  6.2220044  0.8099666  7.6818 1.569e-14 ***
## concen       0.2089935  0.0265297  7.8777 3.334e-15 ***
## ldist       -0.8520919  0.2464836 -3.4570 0.0005462 ***
## I(ldist^2)   0.0974604  0.0186358  5.2297 1.698e-07 ***
## y98          0.0224743  0.0044544  5.0454 4.526e-07 ***
## y99          0.0366898  0.0044528  8.2398 < 2.2e-16 ***
## y00          0.0982120  0.0044576 22.0324 < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Total Sum of Squares:    67.626
## Residual Sum of Squares: 52.162
## R-Squared:              0.22866
## Adj. R-Squared: 0.22766
## Chisq: 1360.42 on 6 DF, p-value: < 2.22e-16

```

```
stargazer(pooling, between, fixed , random, type = "text")
```

```

##
## =====
=====
##                               Dependent variable:
##                               -----
-----
##                               lfare
##                               (1)          (2)          (3)
(4)
## -----
-----
## concen          0.360***          0.382***          0.169***          0.20
9***
##              (0.030)          (0.061)          (0.029)          (0.0
27)
##
## ldist          -0.902***          -0.909***          -0.85
2***
##              (0.128)          (0.247)          (0.2
46)
##
## I(ldist2)       0.103***          0.104***          0.09
7***
##              (0.010)          (0.019)          (0.0
19)
##
## y98            0.021            0.023***          0.02
2***
##              (0.014)            (0.004)          (0.0
04)
##
## y99            0.038***            0.036***          0.03
7***
##              (0.014)            (0.004)          (0.0
04)
##
## y00            0.100***            0.098***          0.09
8***
##              (0.014)            (0.004)          (0.0
04)
##
## Constant       6.209***          6.247***          6.22
2***
##              (0.421)          (0.809)          (0.8
10)
##
## -----
-----
## Observations    4,596            1,149            4,596            4,5
96
## R2              0.406            0.422            0.135            0.2
29
## Adjusted R2     0.405            0.420            -0.154            0.2
28
## F Statistic    523.175*** (df = 6; 4589) 278.178*** (df = 3; 1145) 134.611*** (df = 4; 3443) 1,360.
422***
## =====
=====
## Note:
p<0.01


*p<0.1; **p<0.05; ***


```

# 11 DIAGNOSIS de los modelos: ¿qué método de regresión debería utilizar?

## 11.1 ¿Existen efectos individuales? (paneles heterogenos)

En primer lugar tengo que contrastar la existencia de efectos fijos individuales.

Cuando tengo grados de libertad suficiente es posible contrastar si todas las pendientes son iguales o no

```
# En nuestro caso no podemos ejecutarlo porque da error
# mvariable<-pvcml(fare ~ concen + ldist + I(ldist^2) + y98 + y99 + y00, data=pdata, model= "within")
# fixed <- plm(fare ~ concen + ldist + I(ldist^2) + y98 + y99 + y00, data=pdata, model= "within")
# pooltest(fixed,mvariable)

# Alternativamente podemos hacer un contraste de la existencia de efectos fijos heterogeneos
# LM test for fixed effects versus OLS
plmtest(pooling)
```

```
##
## Lagrange Multiplier Test - (Honda) for balanced panels
##
## data: fare ~ concen + ldist + I(ldist^2) + y98 + y99 + y00
## normal = 74.606, p-value < 2.2e-16
## alternative hypothesis: significant effects
```

```
head(summary(fixef(fixed, type="level")),20)
```

```
##      Estimate Std. Error   t-value Pr(>|t|)
## 1  4.536398  0.05873607  77.23360      0
## 2  4.583861  0.05626634  81.46720      0
## 3  5.206069  0.05815075  89.52710      0
## 4  5.267308  0.05483261  96.06159      0
## 5  4.604657  0.05541569  83.09302      0
## 6  5.021535  0.05994560  83.76821      0
## 7  5.225488  0.05685448  91.90987      0
## 8  4.854573  0.05627291  86.26838      0
## 9  4.961961  0.05375440  92.30800      0
## 10 5.378720  0.05393020  99.73483      0
## 11 5.095338  0.05502583  92.59903      0
## 12 4.529817  0.05687312  79.64778      0
## 13 4.988824  0.05812147  85.83445      0
## 14 4.091223  0.05856557  69.85714      0
## 15 4.832458  0.05525124  87.46334      0
## 16 4.776624  0.05719860  83.50947      0
## 17 4.434304  0.05919572  74.90920      0
## 18 4.572558  0.05882454  77.73215      0
## 19 5.113569  0.05657635  90.38351      0
## 20 5.414561  0.05389760 100.46015      0
```

```
pFtest(fixed, pooling)
```

```
##  
## F test for individual effects  
##  
## data: lfare ~ concen + ldist + I(ldist^2) + y98 + y99 + y00  
## F = 36.965, df1 = 1146, df2 = 3443, p-value < 2.2e-16  
## alternative hypothesis: significant effects
```

## 11.2 ¿Tengo un problema de endogeneidad?

```
# Hausman test for fixed versus random effects model  
phtest(random, fixed)
```

```
##  
## Hausman Test  
##  
## data: lfare ~ concen + ldist + I(ldist^2) + y98 + y99 + y00  
## chisq = 9.9967, df = 4, p-value = 0.04048  
## alternative hypothesis: one model is inconsistent
```

## 12 El problema de la heterocedasticidad entre individuos

Hasta hora, suponíamos que todos los errores de todos los individuos tienen la misma varianza. Tratándose de datos de diferentes individuos habría que asegurarse de ese supuesto.

### 12.1 Estimación Robusta de las Varianzas de los estimadores

```
#¿qué pasa con la heteroscedasticidad?: estimación robusta de las VAR(betas)  
fixed <- plm(lfare ~ concen + ldist + I(ldist^2) + y98 + y99 + y00, data=pdata, model= "within")  
summary(fixed)
```

```
## Oneway (individual) effect Within Model
##
## Call:
## plm(formula = lfare ~ concen + ldist + I(ldist^2) + y98 + y99 +
##      y00, data = pdata, model = "within")
##
## Balanced Panel: n = 1149, T = 4, N = 4596
##
## Residuals:
##      Min.      1st Qu.      Median      3rd Qu.      Max.
## -0.8812810 -0.0397341  0.0023907  0.0436350  0.9188102
##
## Coefficients:
##      Estimate Std. Error t-value Pr(>|t|)
## concen 0.1688590  0.0294101  5.7415 1.020e-08 ***
## y98     0.0228328  0.0044515  5.1292 3.071e-07 ***
## y99     0.0363819  0.0044495  8.1766 4.061e-16 ***
## y00     0.0977717  0.0044555 21.9441 < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Total Sum of Squares:    45.169
## Residual Sum of Squares: 39.06
## R-Squared:    0.13524
## Adj. R-Squared: -0.1541
## F-statistic: 134.611 on 4 and 3443 DF, p-value: < 2.22e-16
```

```
suppressMessages(library(lmtest))
suppressMessages(library(sandwich))

# Para la estimación robusta de la matriz de Var.Cov de los estimadores utilizamos la librería lmtes
# t y sandwich

fixed.ro<-coeftest(fixed, vcov = vcovHC(fixed, type = "HC1",cluster = "group",adjust = T))

# Da el mismo resultado que fixed.ro<-coeftest(fixed, vcov = vcovHC(fixed, type = "HC1"))

random.ro<-coeftest(random, vcov = vcovHC(fixed, type = "HC1",cluster = "group",adjust = T))

# para sacar el mismo resultado que stata hay que ajustar los grados de libertad
# numid<-length(unique(pdata$id))
# mT<-length(pdata$id)
# dfa<-(numid/(numid-1))*(mT-1)/fixed$df.residual
#
# fixed.rostata<-coeftest(fixed, vcov = dfa*vcovHC(fixed, type = "HC1",cluster = "group",adjust =
# T))
# random.rostata<-coeftest(random, vcov = vcovHC(random, type = "HC1",cluster = "group",adjust = T))

stargazer(pooling, between, fixed.ro, random.ro, type = "text")
```

```
##
## =====
##                               Dependent variable:
##                               -----
##                               lfare
##                               panel
##                               linear
##                               coefficient
##                               test
##                               (1)          (2)          (3)          (4)
## -----
## concen          0.360***          0.382***          0.169*** 0.209***
##                  (0.030)          (0.061)          (0.049) (0.049)
##
## ldist           -0.902***          -0.909***
##                  (0.128)          (0.247)
##
## I(ldist2)        0.103***          0.104***
##                  (0.010)          (0.019)
##
## y98              0.021              0.023*** 0.022***
##                  (0.014)              (0.004) (0.004)
##
## y99              0.038***              0.036*** 0.037***
##                  (0.014)              (0.005) (0.005)
##
## y00              0.100***              0.098*** 0.098***
##                  (0.014)              (0.006) (0.006)
##
## Constant         6.209***          6.247***
##                  (0.421)          (0.809)
## -----
## Observations      4,596              1,149
## R2                 0.406              0.422
## Adjusted R2       0.405              0.420
## F Statistic 523.175*** (df = 6; 4589) 278.178*** (df = 3; 1145)
## =====
## Note:                                     *p<0.1; **p<0.05; ***p<0.01
```

## 12.2 Estimación de efectos fijos por mínimos Cuadrados Generalizados (Cluster de individuos)

Es posible estimar el modelo de efectos fijos pero permitiendo que las varianzas entre individuos diferentes sea diferente y estimando por mínimos cuadrados Generalizados Sólo necesario para el estimador de efectos Fijos (efectos aleatorios ya considera varianzas diferentes)

```
fixed.ro2 <- pgglsl(lfare ~ concen + ldist + I(ldist^2) + y98 + y99 + y00, data=pdata, model= "withi
n")
summary(fixed.ro2)
```

```
## Oneway (individual) effect Within FGLS model
##
## Call:
## pgls(formula = lfare ~ concen + ldist + I(ldist^2) + y98 + y99 +
##       y00, data = pdata, model = "within")
##
## Balanced Panel: n = 1149, T = 4, N = 4596
##
## Residuals:
##      Min.      1st Qu.      Median      Mean      3rd Qu.      Max.
## -0.881203 -0.039694  0.002317  0.000000  0.043574  0.919670
##
## Coefficients:
##      Estimate Std. Error z-value Pr(>|z|)
## concen 0.1445250  0.0282987  5.1071 3.271e-07 ***
## y98     0.0230501  0.0040782  5.6520 1.586e-08 ***
## y99     0.0361951  0.0052072  6.9510 3.627e-12 ***
## y00     0.0975047  0.0056114 17.3762 < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## Total Sum of Squares: 875.09
## Residual Sum of Squares: 39.068
## Multiple R-squared: 0.95536
```

## 13 ¿Estimador CRE Se podría obtener de alguna manera la estimación de los efectos fijos individuales si observables?

Al estimar por efectos fijos (cuando existe un problema de endogeneidad entre los efectos fijos individuales no observables y alguna variable explicativa) Desaparecen del modelo la estimación de todos los efectos fijos individuales, tanto los observables como los no observables.

En nuestro ejemplo ¿podría estimarse mediante el estimador de efectos fijos el efecto de la distancia del vuelo (que no varía en el tiempo)?

En principio existen dos alternativas

La Primera alternativa es obtener los efectos fijos y después hacer una segunda regresión de estos efectos fijos sobre los efectos fijos observables

```
efijos<-data.frame(ldist=between(pdata$ldist), ldist2=between(pdata$ldist^2),ef=as.numeric(fixef(fixed, type="level")))
efijos.lm<-lm(ef~ldist + ldist2, data=efijos)
summary(efijos.lm)
```

```
##
## Call:
## lm(formula = ef ~ ldist + ldist2, data = efijos)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.69697 -0.25555 -0.02697  0.23199  0.89593
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  6.22539    0.81312   7.656 4.05e-14 ***
## ldist        -0.83894    0.24729  -3.393 0.000716 ***
## ldist2        0.09598    0.01868   5.138 3.27e-07 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.3253 on 1146 degrees of freedom
## Multiple R-squared:  0.4388, Adjusted R-squared:  0.4378
## F-statistic:  448 on 2 and 1146 DF,  p-value: < 2.2e-16
```

La segunda alternativa es utilizar el estimador **CRE** Correlated Random Effects ## Estimador **CRE**

```
stargazer(pooling, between, fixed, random, type = "text")
```



```

##
## =====
=====
##                               Dependent variable:
##                               -----
-----
##                               lfare
##                               (1)          (2)          (3)
(4)
## -----
-----
## concen          0.360***          0.382***          0.169***          0.20
9***
##              (0.030)          (0.061)          (0.029)          (0.0
27)
##
## ldist          -0.902***          -0.909***          -0.85
2***
##              (0.128)          (0.247)          (0.2
46)
##
## I(ldist2)       0.103***          0.104***          0.09
7***
##              (0.010)          (0.019)          (0.0
19)
##
## y98            0.021            0.023***          0.02
2***
##              (0.014)            (0.004)          (0.0
04)
##
## y99            0.038***            0.036***          0.03
7***
##              (0.014)            (0.004)          (0.0
04)
##
## y00            0.100***            0.098***          0.09
8***
##              (0.014)            (0.004)          (0.0
04)
##
## Constant       6.209***          6.247***          6.22
2***
##              (0.421)          (0.809)          (0.8
10)
##
## -----
-----
## Observations    4,596            1,149            4,596            4,5
96
## R2              0.406            0.422            0.135            0.2
29
## Adjusted R2     0.405            0.420            -0.154            0.2
28
## F Statistic    523.175*** (df = 6; 4589) 278.178*** (df = 3; 1145) 134.611*** (df = 4; 3443) 1,360.
422***
## =====
=====
## Note:
p<0.01


*p<0.1; **p<0.05; ***


```

Observesen la tabla anterior que como la distancia de un mismo vuelo (ej NuevaYork-Londres) no varía en el tiempo, esta variable distancia sale del modelo de efectos fijos.

Se puede obtener la estimación de *Efectos Fijos* de estas variables individuales fijas que sí son observables utilizando el estimador **CRE**

El método **CRE** consiste en lo siguiente:

1. Se hace la media de cada grupo de las variables que varían en  $i,t$  (ni las variables fijas ni los efectos temporales se incluyen). puede hacerse por ejemplo con `Between` en mayúsculas

```
pdata$concen_bar<-Between(pdata$concen)
```

2. Y ahora se incluye estas medias en la regresión pero con efectos aleatorios: **El resultado es el estimador de efectos fijos** para esas variables fijas sí observables (observese que para el resto de variables se obtiene exactamente la misma estimación que con efectos fijos)

```
CRE <- plm(lfare ~ concen + ldist + I(ldist^2) + y98 + y99 + y00 + concen_bar, data=pdata, model= "random")
summary(CRE)
```

```
## Oneway (individual) effect Random Effect Model
##      (Swamy-Arora's transformation)
##
## Call:
## plm(formula = lfare ~ concen + ldist + I(ldist^2) + y98 + y99 +
##      y00 + concen_bar, data = pdata, model = "random")
##
## Balanced Panel: n = 1149, T = 4, N = 4596
##
## Effects:
##              var std.dev share
## idiosyncratic 0.01134 0.10651  0.1
## individual    0.10198 0.31934  0.9
## theta: 0.8355
##
## Residuals:
##      Min.      1st Qu.      Median      3rd Qu.      Max.
## -0.8363877 -0.0636849 -0.0039541  0.0628812  0.8598019
##
## Coefficients:
##              Estimate Std. Error z-value Pr(>|z|)
## (Intercept)  6.2078886  0.8091936  7.6717 1.697e-14 ***
## concen       0.1688590  0.0294101  5.7415 9.383e-09 ***
## ldist       -0.9089296  0.2469054 -3.6813 0.0002321 ***
## I(ldist^2)   0.1038426  0.0187278  5.5448 2.942e-08 ***
## y98          0.0228328  0.0044515  5.1292 2.910e-07 ***
## y99          0.0363819  0.0044495  8.1766 2.920e-16 ***
## y00          0.0977717  0.0044555 21.9441 < 2.2e-16 ***
## concen_bar   0.2136346  0.0678537  3.1485 0.0016413 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Total Sum of Squares:    67.626
## Residual Sum of Squares: 52.05
## R-Squared:              0.23033
## Adj. R-Squared:         0.22915
## Chisq: 1372.98 on 7 DF, p-value: < 2.22e-16
```

y si tengo un problema de heterocedasticidad entre individuos, estimo la varianz de los estimadores de forma robusta

```
stargazer(fixed, random, CRE, type = "text")
```

##	=====			
##	Dependent variable:			
##	-----			
##	lfare			
##	(1)	(2)	(3)	
##	-----			
##	concen	0.169***	0.209***	0.169***
##		(0.029)	(0.027)	(0.029)
##				
##	ldist	-0.852***	-0.909***	
##		(0.246)	(0.247)	
##				
##	I(ldist2)	0.097***	0.104***	
##		(0.019)	(0.019)	
##				
##	y98	0.023***	0.022***	0.023***
##		(0.004)	(0.004)	(0.004)
##				
##	y99	0.036***	0.037***	0.036***
##		(0.004)	(0.004)	(0.004)
##				
##	y00	0.098***	0.098***	0.098***
##		(0.004)	(0.004)	(0.004)
##				
##	concen_bar			0.214***
##				(0.068)
##				
##	Constant		6.222***	6.208***
##			(0.810)	(0.809)
##				
##	-----			
##	Observations	4,596	4,596	4,596
##	R2	0.135	0.229	0.230
##	Adjusted R2	-0.154	0.228	0.229
##	F Statistic	134.611*** (df = 4; 3443)	1,360.422***	1,372.977***
##	=====			
##	Note: *p<0.1; **p<0.05; ***p<0.01			



Proyecto de innovación  
“Innova Docencia” de la UCM  
Convocatoria 2018-2019  
Proyecto número 275

**R-adaptación de la asignatura de métodos econométricos en economía y finanzas del  
Grado de Estadística Aplicada**

Facultad de Estudios Estadísticos  
Departamento de Economía Aplicada, Pública y Política

**Título**

***Práctica 17: La ecuación de Salarios con Datos de Panel***

**Correspondiente al**

**Capítulo 8 MODELOS DE REGRESIÓN CON DATOS DE PANEL**

- Especificación de modelos con datos de panel
  - Estimación de modelos con efectos fijos
  - Estimación de modelos con efectos aleatorios
  - Sistema de ecuaciones aparentemente no relacionadas
  - Estimador de efectos fijos observables con el estimador CRE
- Práctica 16: Modelo de datos de Panel con R: la librería “pml”***  
***Práctica 17: La ecuación de Salarios con Datos de Panel***

# Modelos de regresión con Datos de Panel en R: Ejemplo sobre la ecuación de salarios

## Asignatura Métodos Econométricos en Economía y Finanzas

Profesores  
Loren Escot, Alicia Pérez-Alonso y Julio E. Sandubete

Dpto Economía Aplicada, Pública y Política  
Facultad Estudios Estadísticos - Universidad Complutense de Madrid

Curso académico 2018/2019

### Abstract

En esta práctica se revisa la estimación de modelos de regresión con Datos de Panel. Se presentará la librería `plm` para trabajar con datos de panel, esto es, con datos que tienen una dimensión temporal y otra de sección cruzada. Se presentarán los diferentes métodos de regresión como el modelo pooling, between, Fixed effects y random effects, y se presentarán los diferentes contrastes para poder saber cual de ellos es mejor. También presentamos la estimación CRE de efectos fijos para estimar efectos fijos si observables.

Se presentarán las diferentes técnicas utilizando un ejemplo de utilizado en el manual de Wooldridge. trabajaremos con un conjunto de datos panel de individuos a lo largo de 7 años 1980 a 1987 para estimar una ecuación de salarios. Estimaremos el efecto de la educación solucionando el problema de la endogeneidad en caso de que exista

- 1 Definimos la estructura de los datos de Panel
- 2 Estadística descriptiva
- 3 Pooled OLS estimator
- 4 Between estimator
- 5 Efectos fijos y primeras diferencias
  - 5.1 Estimador de primeras diferencias
  - 5.2 Estimador de efectos Fijos
- 6 Estimador de Efectos Aleatorios
- 7 Diagnóstico del modelo
- 8 Corrección por heterocedasticidad
- 9 Estimador CRE
  - 9.1 Correlated Random Effects

```
#=====
```

```
# Datos de panel en R
```

```
# '# Cargo librerías necesarias
```

```
suppressMessages(library(stargazer))
```

```
suppressMessages(library(wooldridge))
```

```
suppressMessages(library(plm)) # librería para estimar con datos de panel
```

```
# '# Cargo los datos
```

```
# Ejemplo 14.4 de la pag 443 del libro de wooldridge
```

```
data("wagepan")
```

```
#?wagepan
```

```
# trabajaremos con un conjunto de datos panel de individuos a lo largo de 7 años 1980 a 1987
```

```
# para estimar una ecuación de salarios. Estimaremos el efecto de la educación
```

```
# solucionando el problema de la endogeneidad en caso de que exista
```

```
# En principio seguiremos el ejemplo 14.4 del libro de Wooldridge
```

```
#el modelo será
```

```
modelo<- lwage~ educ + black + hisp + exper + exper2 + union + married
```

```
summary(wagepan)
```

##	nr	year	agric	black
##	Min. : 13	Min. :1980	Min. :0.00000	Min. :0.0000
##	1st Qu.: 2329	1st Qu.:1982	1st Qu.:0.00000	1st Qu.:0.0000
##	Median : 4569	Median :1984	Median :0.00000	Median :0.0000
##	Mean : 5262	Mean :1984	Mean :0.03211	Mean :0.1156
##	3rd Qu.: 8406	3rd Qu.:1985	3rd Qu.:0.00000	3rd Qu.:0.0000
##	Max. :12548	Max. :1987	Max. :1.00000	Max. :1.0000
##	bus	construc	ent	exper
##	Min. :0.00000	Min. :0.000	Min. :0.00000	Min. : 0.000
##	1st Qu.:0.00000	1st Qu.:0.000	1st Qu.:0.00000	1st Qu.: 4.000
##	Median :0.00000	Median :0.000	Median :0.00000	Median : 6.000
##	Mean :0.07592	Mean :0.075	Mean :0.01514	Mean : 6.515
##	3rd Qu.:0.00000	3rd Qu.:0.000	3rd Qu.:0.00000	3rd Qu.: 9.000
##	Max. :1.00000	Max. :1.000	Max. :1.00000	Max. :18.000
##	fin	hisp	poorhlth	hours
##	Min. :0.00000	Min. :0.000	Min. :0.00000	Min. : 120
##	1st Qu.:0.00000	1st Qu.:0.000	1st Qu.:0.00000	1st Qu.:2040
##	Median :0.00000	Median :0.000	Median :0.00000	Median :2080
##	Mean :0.03693	Mean :0.156	Mean :0.01697	Mean :2191
##	3rd Qu.:0.00000	3rd Qu.:0.000	3rd Qu.:0.00000	3rd Qu.:2414
##	Max. :1.00000	Max. :1.000	Max. :1.00000	Max. :4992
##	manuf	married	min	nrthcen
##	Min. :0.0000	Min. :0.000	Min. :0.0000	Min. :0.0000
##	1st Qu.:0.0000	1st Qu.:0.000	1st Qu.:0.0000	1st Qu.:0.0000
##	Median :0.0000	Median :0.000	Median :0.0000	Median :0.0000
##	Mean :0.2823	Mean :0.439	Mean :0.0156	Mean :0.2578
##	3rd Qu.:1.0000	3rd Qu.:1.000	3rd Qu.:0.0000	3rd Qu.:1.0000
##	Max. :1.0000	Max. :1.000	Max. :1.0000	Max. :1.0000
##	nrtheast	occ1	occ2	occ3
##	Min. :0.0000	Min. :0.0000	Min. :0.00000	Min. :0.00000
##	1st Qu.:0.0000	1st Qu.:0.0000	1st Qu.:0.00000	1st Qu.:0.00000
##	Median :0.0000	Median :0.0000	Median :0.00000	Median :0.00000
##	Mean :0.1901	Mean :0.1039	Mean :0.09151	Mean :0.05344
##	3rd Qu.:0.0000	3rd Qu.:0.0000	3rd Qu.:0.00000	3rd Qu.:0.00000
##	Max. :1.0000	Max. :1.0000	Max. :1.00000	Max. :1.00000
##	occ4	occ5	occ6	occ7
##	Min. :0.0000	Min. :0.0000	Min. :0.0000	Min. :0.00000
##	1st Qu.:0.0000	1st Qu.:0.0000	1st Qu.:0.0000	1st Qu.:0.00000
##	Median :0.0000	Median :0.0000	Median :0.0000	Median :0.00000
##	Mean :0.1115	Mean :0.2142	Mean :0.2021	Mean :0.09197
##	3rd Qu.:0.0000	3rd Qu.:0.0000	3rd Qu.:0.0000	3rd Qu.:0.00000
##	Max. :1.0000	Max. :1.0000	Max. :1.0000	Max. :1.00000
##	occ8	occ9	per	pro
##	Min. :0.00000	Min. :0.0000	Min. :0.00000	Min. :0.00000
##	1st Qu.:0.00000	1st Qu.:0.0000	1st Qu.:0.00000	1st Qu.:0.00000
##	Median :0.00000	Median :0.0000	Median :0.00000	Median :0.00000
##	Mean :0.01468	Mean :0.1167	Mean :0.01674	Mean :0.07638
##	3rd Qu.:0.00000	3rd Qu.:0.0000	3rd Qu.:0.00000	3rd Qu.:0.00000
##	Max. :1.00000	Max. :1.0000	Max. :1.00000	Max. :1.00000
##	pub	rur	south	educ
##	Min. :0.00000	Min. :0.0000	Min. :0.0000	Min. : 3.00
##	1st Qu.:0.00000	1st Qu.:0.0000	1st Qu.:0.0000	1st Qu.:11.00
##	Median :0.00000	Median :0.0000	Median :0.0000	Median :12.00
##	Mean :0.04014	Mean :0.2039	Mean :0.3507	Mean :11.77
##	3rd Qu.:0.00000	3rd Qu.:0.0000	3rd Qu.:1.0000	3rd Qu.:12.00
##	Max. :1.00000	Max. :1.0000	Max. :1.0000	Max. :16.00
##	tra	trad	union	lwage
##	Min. :0.0000	Min. :0.0000	Min. :0.000	Min. : -3.579
##	1st Qu.:0.0000	1st Qu.:0.0000	1st Qu.:0.000	1st Qu.: 1.351
##	Median :0.0000	Median :0.0000	Median :0.000	Median : 1.671
##	Mean :0.0656	Mean :0.2681	Mean :0.244	Mean : 1.649

```
## 3rd Qu.:0.0000 3rd Qu.:1.0000 3rd Qu.:0.000 3rd Qu.: 1.991
## Max. :1.0000 Max. :1.0000 Max. :1.000 Max. : 4.052
## d81 d82 d83 d84
## Min. :0.000 Min. :0.000 Min. :0.000 Min. :0.000
## 1st Qu.:0.000 1st Qu.:0.000 1st Qu.:0.000 1st Qu.:0.000
## Median :0.000 Median :0.000 Median :0.000 Median :0.000
## Mean :0.125 Mean :0.125 Mean :0.125 Mean :0.125
## 3rd Qu.:0.000 3rd Qu.:0.000 3rd Qu.:0.000 3rd Qu.:0.000
## Max. :1.000 Max. :1.000 Max. :1.000 Max. :1.000
## d85 d86 d87 expersq
## Min. :0.000 Min. :0.000 Min. :0.000 Min. : 0.00
## 1st Qu.:0.000 1st Qu.:0.000 1st Qu.:0.000 1st Qu.: 16.00
## Median :0.000 Median :0.000 Median :0.000 Median : 36.00
## Mean :0.125 Mean :0.125 Mean :0.125 Mean : 50.42
## 3rd Qu.:0.000 3rd Qu.:0.000 3rd Qu.:0.000 3rd Qu.: 81.00
## Max. :1.000 Max. :1.000 Max. :1.000 Max. :324.00
```

```
head(wagepan,n = 20)
```

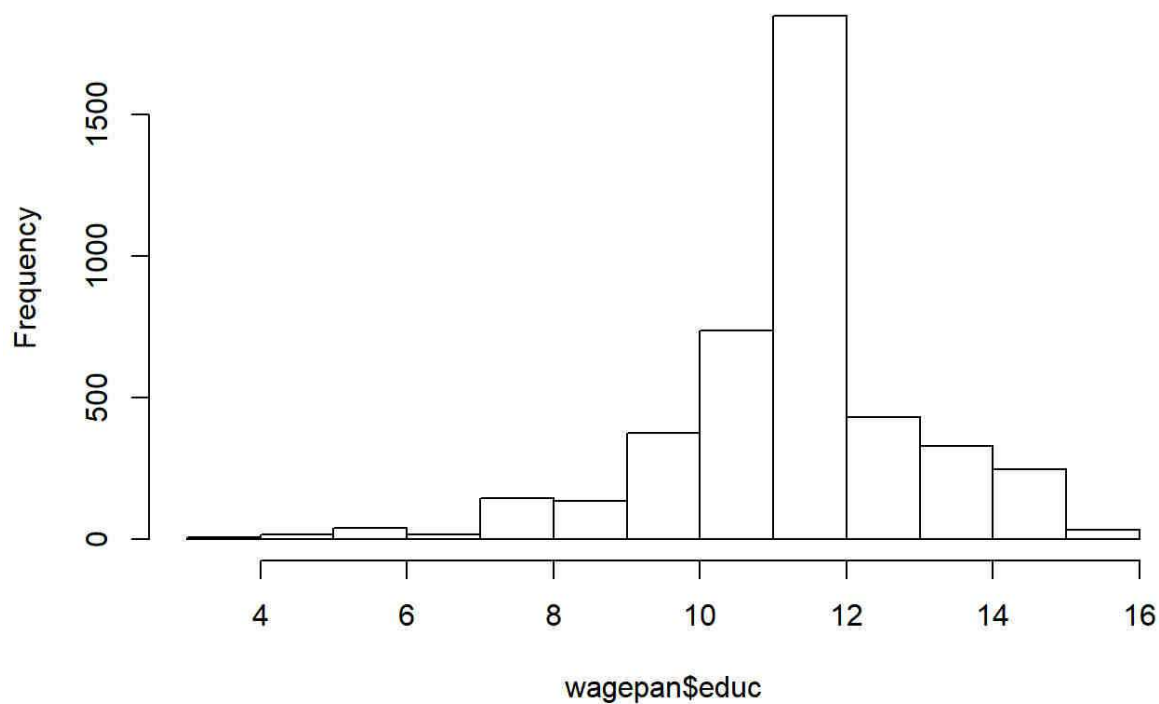


##	nr	year	agric	black	bus	construc	ent	exper	fin	hisp	poorhlth	hours		
## 1	13	1980	0	0	1	0	0	1	0	0	0	2672		
## 2	13	1981	0	0	0	0	0	2	0	0	0	2320		
## 3	13	1982	0	0	1	0	0	3	0	0	0	2940		
## 4	13	1983	0	0	1	0	0	4	0	0	0	2960		
## 5	13	1984	0	0	0	0	0	5	0	0	0	3071		
## 6	13	1985	0	0	1	0	0	6	0	0	0	2864		
## 7	13	1986	0	0	1	0	0	7	0	0	0	2994		
## 8	13	1987	0	0	1	0	0	8	0	0	0	2640		
## 9	17	1980	0	0	0	0	0	4	0	0	0	2484		
## 10	17	1981	0	0	0	0	0	5	0	0	0	2804		
## 11	17	1982	0	0	0	0	0	6	0	0	0	2530		
## 12	17	1983	0	0	0	0	0	7	0	0	0	2340		
## 13	17	1984	0	0	0	0	0	8	0	0	0	2486		
## 14	17	1985	0	0	0	1	0	9	0	0	0	2164		
## 15	17	1986	0	0	0	1	0	10	0	0	0	2749		
## 16	17	1987	0	0	0	1	0	11	0	0	0	2476		
## 17	18	1980	0	0	0	0	0	4	0	0	0	2332		
## 18	18	1981	0	0	0	0	0	5	0	0	0	2116		
## 19	18	1982	0	0	0	0	0	6	0	0	0	2500		
## 20	18	1983	0	0	0	0	0	7	0	0	0	2474		
##	manuf	married	min	nrthcen	nrtheast	occ1	occ2	occ3	occ4	occ5	occ6	occ7		
## 1	0	0	0	0	1	0	0	0	0	0	0	0		
## 2	0	0	0	0	1	0	0	0	0	0	0	0		
## 3	0	0	0	0	1	0	0	0	0	0	0	0		
## 4	0	0	0	0	1	0	0	0	0	0	0	0		
## 5	0	0	0	0	1	0	0	0	0	1	0	0		
## 6	0	0	0	0	1	0	1	0	0	0	0	0		
## 7	0	0	0	0	1	0	1	0	0	0	0	0		
## 8	0	0	0	0	1	0	1	0	0	0	0	0		
## 9	0	0	0	0	1	0	1	0	0	0	0	0		
## 10	0	0	0	0	1	0	1	0	0	0	0	0		
## 11	0	0	0	0	1	0	1	0	0	0	0	0		
## 12	0	0	0	0	1	0	1	0	0	0	0	0		
## 13	0	0	0	0	1	0	1	0	0	0	0	0		
## 14	0	0	0	0	1	0	0	0	0	0	0	1		
## 15	0	0	0	0	1	0	0	0	0	0	0	1		
## 16	0	0	0	0	1	0	0	0	0	1	0	0		
## 17	0	1	0	0	0	0	0	0	1	0	0	0		
## 18	0	1	0	0	0	0	0	0	1	0	0	0		
## 19	1	1	0	0	0	0	0	1	0	0	0	0		
## 20	0	1	0	0	0	0	0	0	1	0	0	0		
##	occ8	occ9	per	pro	pub	rur	south	educ	tra	trad	union	lwage	d81	d82
## 1	0	1	0	0	0	0	0	14	0	0	0	1.1975402	0	0
## 2	0	1	1	0	0	0	0	14	0	0	1	1.8530600	1	0
## 3	0	1	0	0	0	0	0	14	0	0	0	1.3444617	0	1
## 4	0	1	0	0	0	0	0	14	0	0	0	1.4332134	0	0
## 5	0	0	1	0	0	0	0	14	0	0	0	1.5681251	0	0
## 6	0	0	0	0	0	0	0	14	0	0	0	1.6998910	0	0
## 7	0	0	0	0	0	0	0	14	0	0	0	-0.7202626	0	0
## 8	0	0	0	0	0	0	0	14	0	0	0	1.6691879	0	0
## 9	0	0	0	0	0	0	0	13	0	1	0	1.6759624	0	0
## 10	0	0	0	0	0	0	0	13	0	1	0	1.5183982	1	0
## 11	0	0	0	0	0	0	0	13	0	1	0	1.5591905	0	1
## 12	0	0	0	0	0	0	0	13	0	1	0	1.7254101	0	0
## 13	0	0	0	0	0	0	0	13	0	1	0	1.6220223	0	0
## 14	0	0	0	0	0	0	0	13	0	0	0	1.6085882	0	0
## 15	0	0	0	0	0	0	0	13	0	0	0	1.5723854	0	0
## 16	0	0	0	0	0	0	0	13	0	0	0	1.8203338	0	0
## 17	0	0	0	0	0	0	0	12	0	1	0	1.5159627	0	0
## 18	0	0	0	0	0	0	0	12	0	1	0	1.7353791	1	0

```
## 19  0  0  0  0  0  0  0  12  0  0  0  1.6317437  0  1
## 20  0  0  0  0  0  0  0  0  12  0  1  0  1.9982288  0  0
##      d83 d84 d85 d86 d87 expersq
## 1  0  0  0  0  0  1
## 2  0  0  0  0  0  4
## 3  0  0  0  0  0  9
## 4  1  0  0  0  0  16
## 5  0  1  0  0  0  25
## 6  0  0  1  0  0  36
## 7  0  0  0  1  0  49
## 8  0  0  0  0  1  64
## 9  0  0  0  0  0  16
## 10 0  0  0  0  0  25
## 11 0  0  0  0  0  36
## 12 1  0  0  0  0  49
## 13 0  1  0  0  0  64
## 14 0  0  1  0  0  81
## 15 0  0  0  1  0  100
## 16 0  0  0  0  1  121
## 17 0  0  0  0  0  16
## 18 0  0  0  0  0  25
## 19 0  0  0  0  0  36
## 20 1  0  0  0  0  49
```

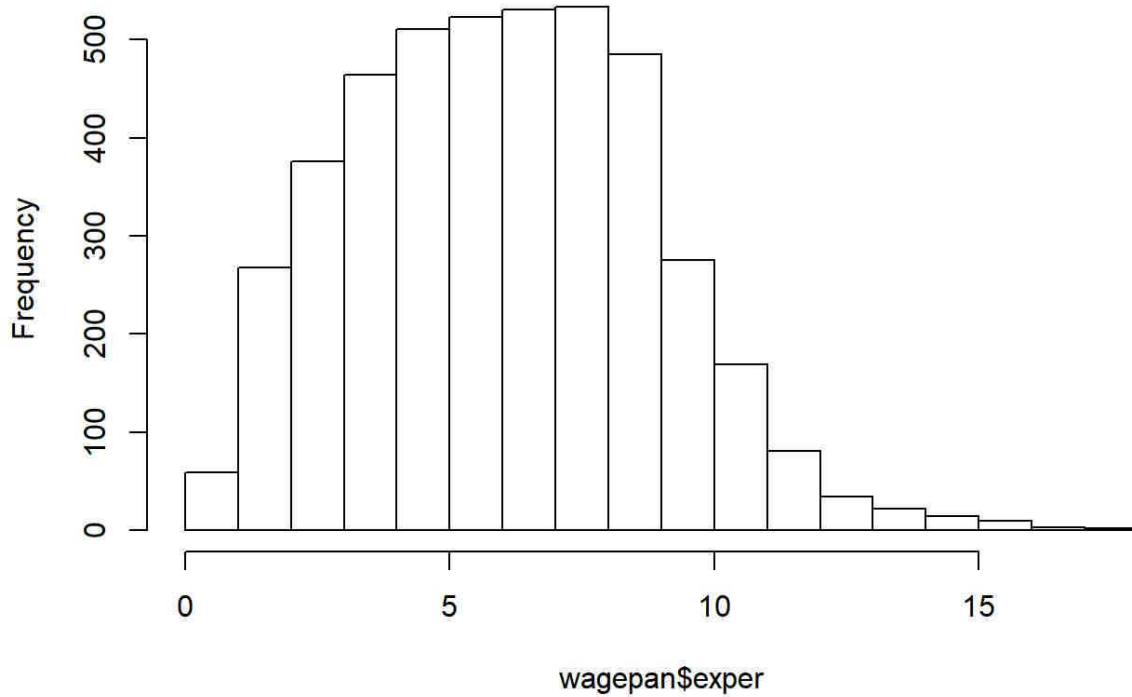
```
#Trabajaremos con las series en logaritmos
hist(wagepan$educ)
```

**Histogram of wagepan\$educ**



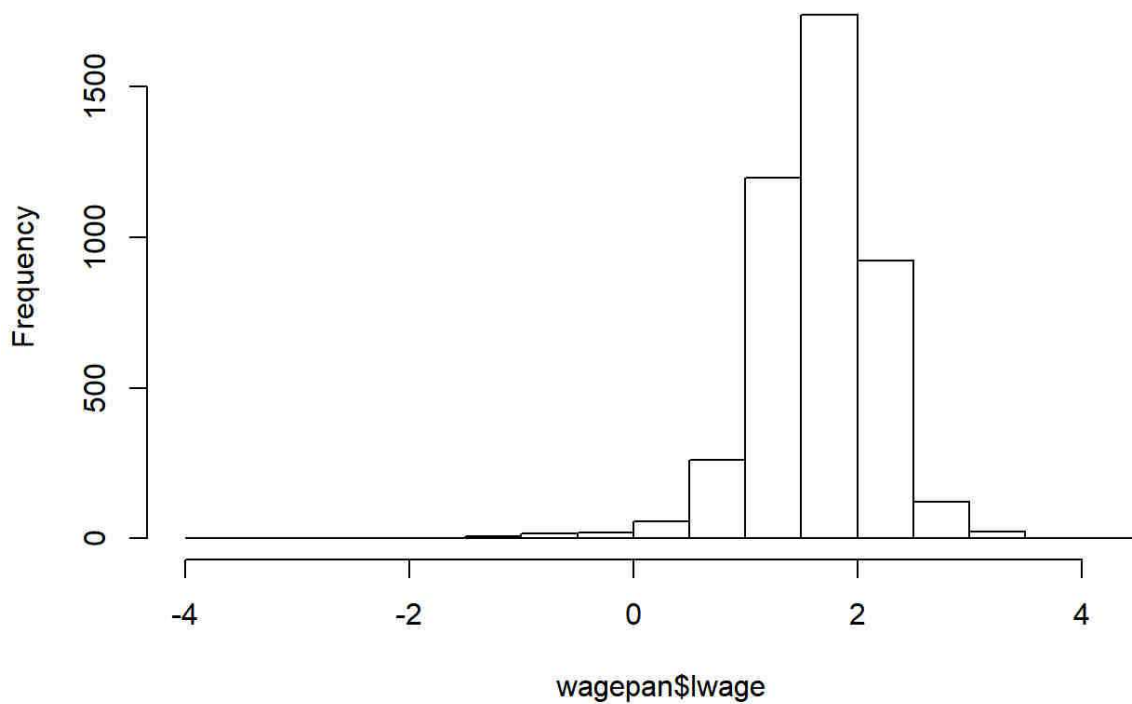
```
hist(wagepan$exper)
```

**Histogram of wagepan\$exper**

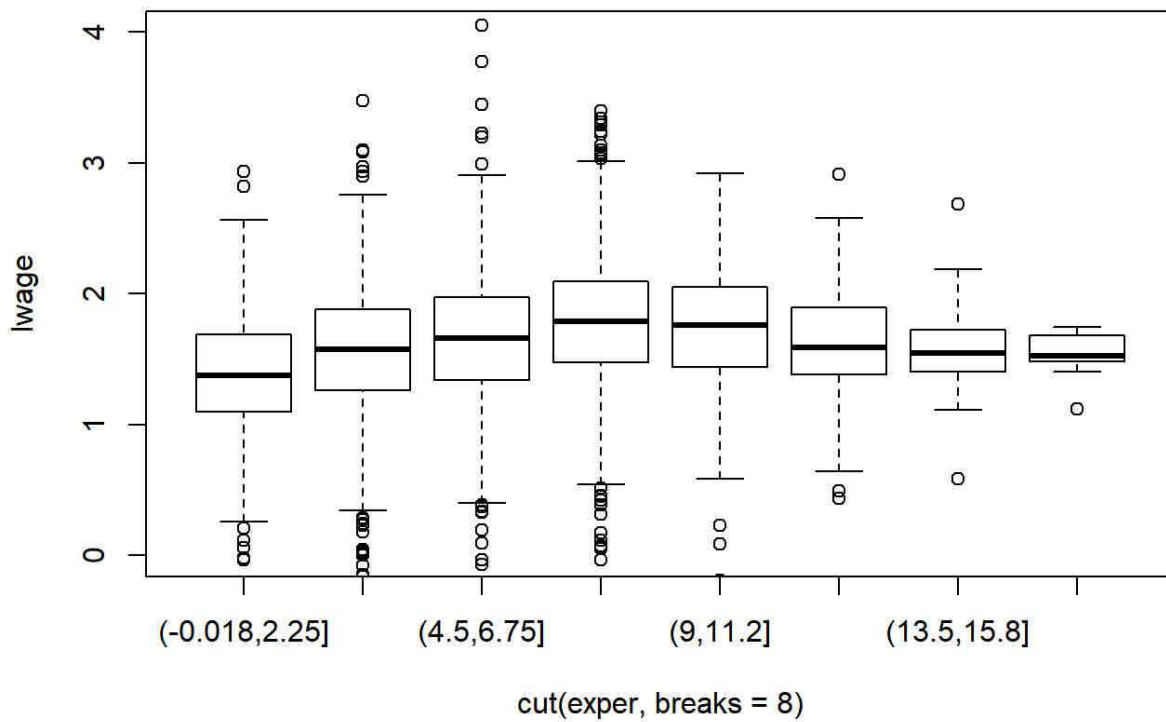


```
hist(wagepan$lwage)
```

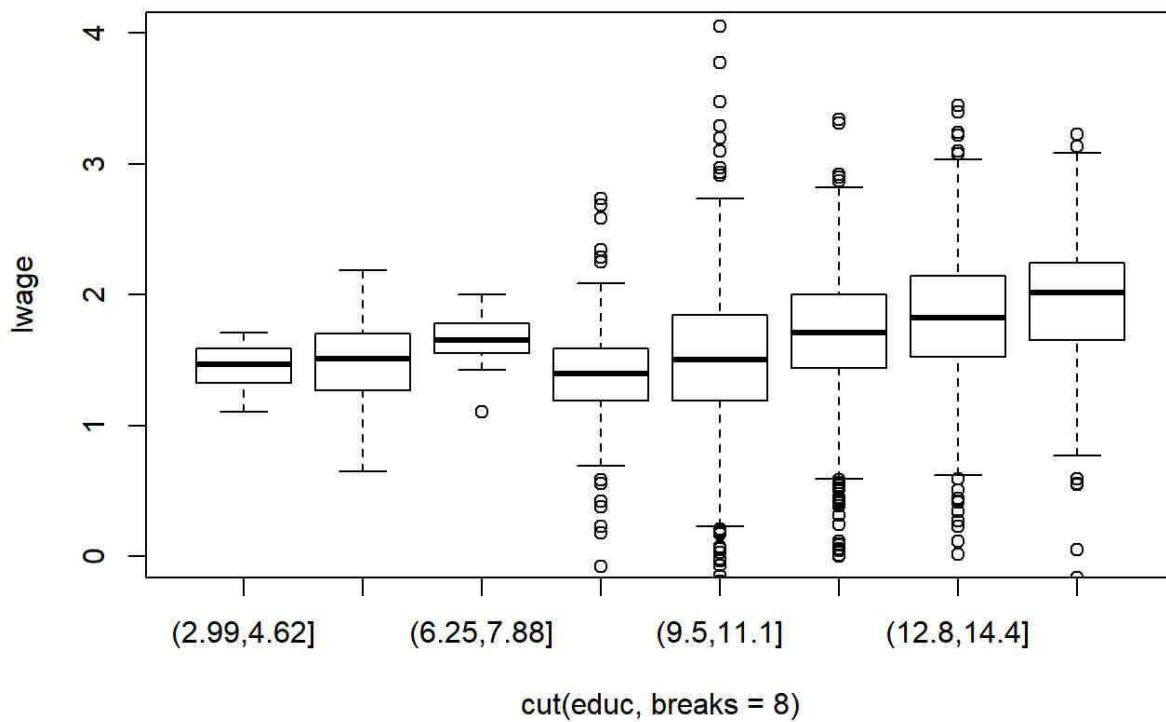
**Histogram of wagepan\$lwage**



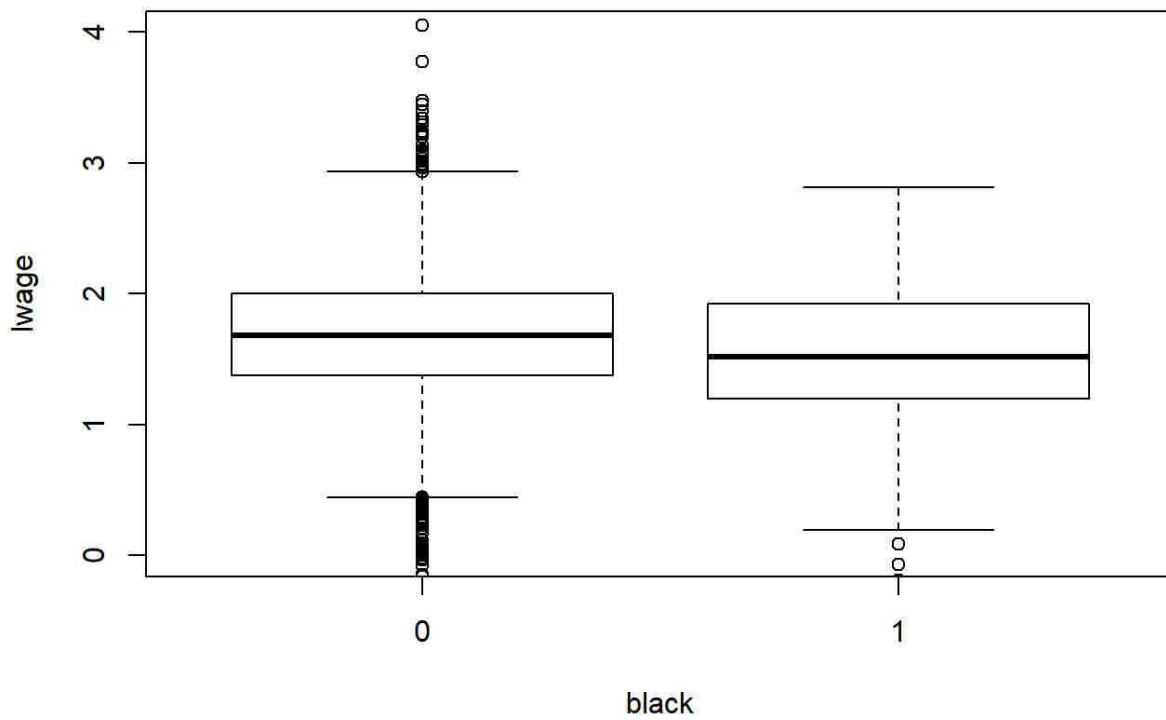
```
boxplot(lwage~cut(exper,breaks = 8),data=wagepan,ylim=c(0,4))
```



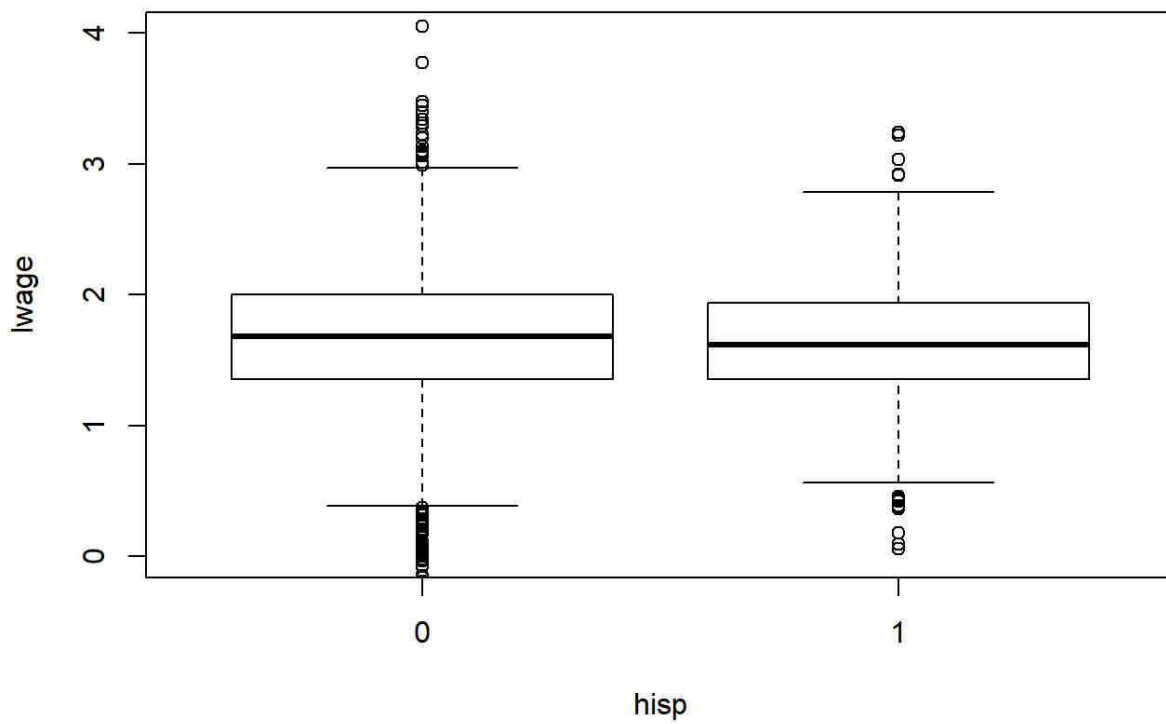
```
boxplot(lwage~cut(exper,breaks = 8),data=wagepan, ylim=c(0,4))
```



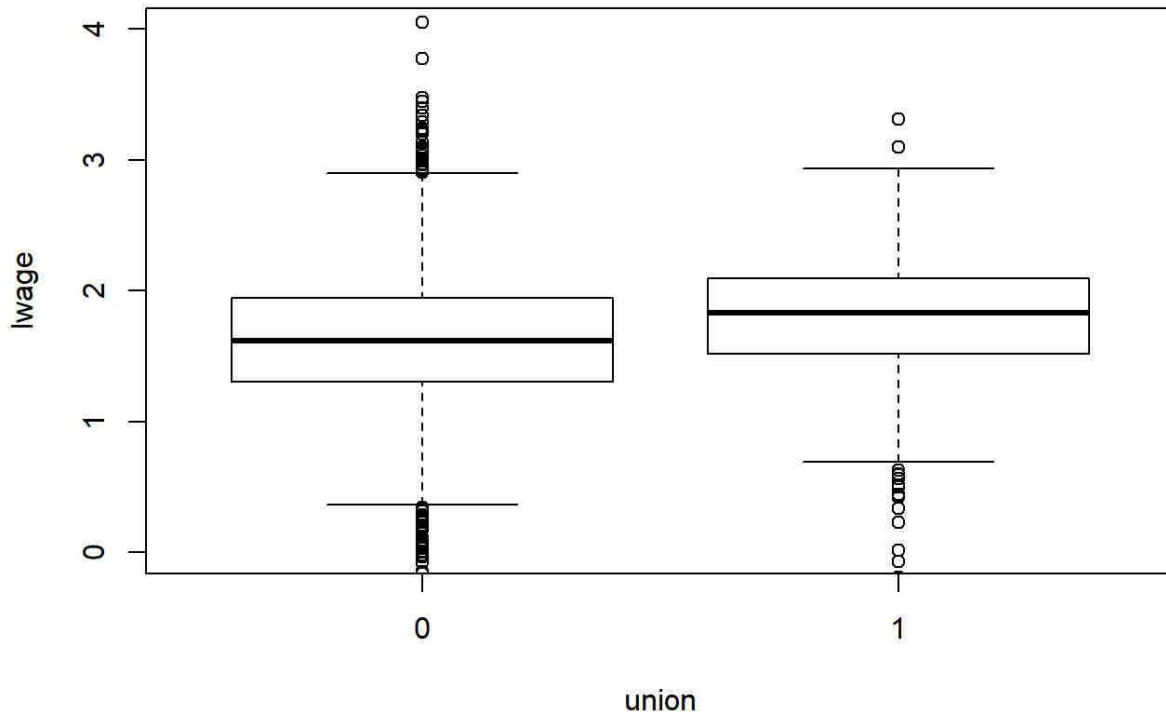
```
boxplot(lwage~black,data=wagepan, ylim=c(0,4))
```



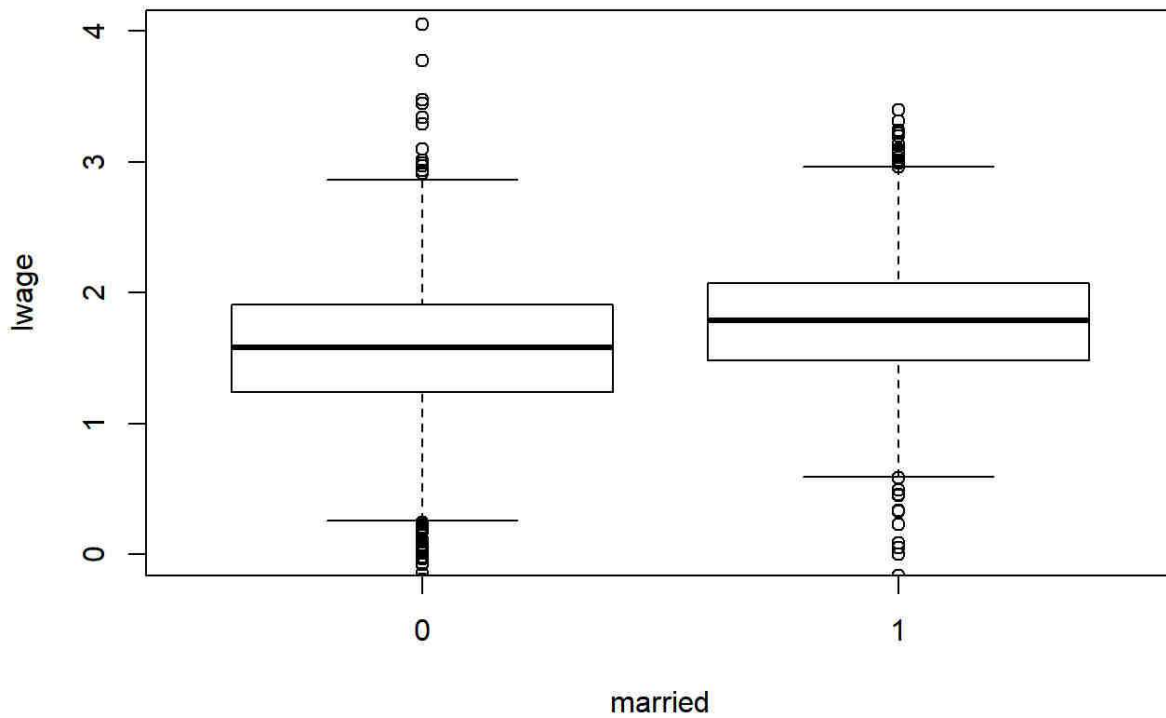
```
boxplot(lwage~hisp,data=wagepan, ylim=c(0,4))
```



```
boxplot(lwage~union,data=wagepan, ylim=c(0,4))
```



```
boxplot(lwage~married,data=wagepan, ylim=c(0,4))
```



1 Definimos la estructura de los datos de Panel

```
# Definimos la estructura de los datos de Panel

# Establecemos los dos indices (de tiempo y de cross-section) para
# que el conjunto de datos sea un panel de datos (objeto pdata.frame)

# Vamos a utilizar el "nr" número de identificación de la persona
# y "year" como identificador del tiempo

# tenemos 8 años
table(wagepan$year)
```

```
##
## 1980 1981 1982 1983 1984 1985 1986 1987
##  545  545  545  545  545  545  545  545
```

```
# y 545 individuos
nrow(table(wagepan$nr))
```

```
## [1] 545
```

```
pdata <- pdata.frame(wagepan, index=c("nr","year"),drop.index=FALSE, row.names = TRUE)
head(pdata,n = 20)
```

##	nr	year	agric	black	bus	construc	ent	exper	fin	hisp	poorhlth	hours
##	13-1980	13 1980	0	0	1	0	0	1	0	0	0	2672
##	13-1981	13 1981	0	0	0	0	0	2	0	0	0	2320
##	13-1982	13 1982	0	0	1	0	0	3	0	0	0	2940
##	13-1983	13 1983	0	0	1	0	0	4	0	0	0	2960
##	13-1984	13 1984	0	0	0	0	0	5	0	0	0	3071
##	13-1985	13 1985	0	0	1	0	0	6	0	0	0	2864
##	13-1986	13 1986	0	0	1	0	0	7	0	0	0	2994
##	13-1987	13 1987	0	0	1	0	0	8	0	0	0	2640
##	17-1980	17 1980	0	0	0	0	0	4	0	0	0	2484
##	17-1981	17 1981	0	0	0	0	0	5	0	0	0	2804
##	17-1982	17 1982	0	0	0	0	0	6	0	0	0	2530
##	17-1983	17 1983	0	0	0	0	0	7	0	0	0	2340
##	17-1984	17 1984	0	0	0	0	0	8	0	0	0	2486
##	17-1985	17 1985	0	0	0	1	0	9	0	0	0	2164
##	17-1986	17 1986	0	0	0	1	0	10	0	0	0	2749
##	17-1987	17 1987	0	0	0	1	0	11	0	0	0	2476
##	18-1980	18 1980	0	0	0	0	0	4	0	0	0	2332
##	18-1981	18 1981	0	0	0	0	0	5	0	0	0	2116
##	18-1982	18 1982	0	0	0	0	0	6	0	0	0	2500
##	18-1983	18 1983	0	0	0	0	0	7	0	0	0	2474
##	manuf	married	min	nrthcen	nrtheast	occ1	occ2	occ3	occ4	occ5	occ6	
##	13-1980	0	0	0	0	1	0	0	0	0	0	0
##	13-1981	0	0	0	0	1	0	0	0	0	0	0
##	13-1982	0	0	0	0	1	0	0	0	0	0	0
##	13-1983	0	0	0	0	1	0	0	0	0	0	0
##	13-1984	0	0	0	0	1	0	0	0	0	1	0
##	13-1985	0	0	0	0	1	0	1	0	0	0	0
##	13-1986	0	0	0	0	1	0	1	0	0	0	0
##	13-1987	0	0	0	0	1	0	1	0	0	0	0
##	17-1980	0	0	0	0	1	0	1	0	0	0	0
##	17-1981	0	0	0	0	1	0	1	0	0	0	0
##	17-1982	0	0	0	0	1	0	1	0	0	0	0
##	17-1983	0	0	0	0	1	0	1	0	0	0	0
##	17-1984	0	0	0	0	1	0	1	0	0	0	0
##	17-1985	0	0	0	0	1	0	0	0	0	0	0
##	17-1986	0	0	0	0	1	0	0	0	0	0	0
##	17-1987	0	0	0	0	1	0	0	0	0	1	0
##	18-1980	0	1	0	0	0	0	0	0	1	0	0
##	18-1981	0	1	0	0	0	0	0	0	1	0	0
##	18-1982	1	1	0	0	0	0	1	0	0	0	0
##	18-1983	0	1	0	0	0	0	0	1	0	0	0
##	occ7	occ8	occ9	per	pro	pub	rur	south	educ	tra	trad	union
##	13-1980	0	0	1	0	0	0	0	14	0	0	0
##	13-1981	0	0	1	1	0	0	0	14	0	0	1
##	13-1982	0	0	1	0	0	0	0	14	0	0	0
##	13-1983	0	0	1	0	0	0	0	14	0	0	0
##	13-1984	0	0	0	1	0	0	0	14	0	0	0
##	13-1985	0	0	0	0	0	0	0	14	0	0	0
##	13-1986	0	0	0	0	0	0	0	14	0	0	0
##	13-1987	0	0	0	0	0	0	0	14	0	0	0
##	17-1980	0	0	0	0	0	0	0	13	0	1	0
##	17-1981	0	0	0	0	0	0	0	13	0	1	0
##	17-1982	0	0	0	0	0	0	0	13	0	1	0
##	17-1983	0	0	0	0	0	0	0	13	0	1	0
##	17-1984	0	0	0	0	0	0	0	13	0	1	0
##	17-1985	1	0	0	0	0	0	0	13	0	0	0
##	17-1986	1	0	0	0	0	0	0	13	0	0	0
##	17-1987	0	0	0	0	0	0	0	13	0	0	0
##	18-1980	0	0	0	0	0	0	0	12	0	1	0
##	18-1981	0	0	0	0	0	0	0	12	0	1	0



## 18-1982	0	0	0	0	0	0	0	0	12	0	0	0
## 18-1983	0	0	0	0	0	0	0	0	12	0	1	0
##	lwage	d81	d82	d83	d84	d85	d86	d87	expersq			
## 13-1980	1.1975402	0	0	0	0	0	0	0	1			
## 13-1981	1.8530600	1	0	0	0	0	0	0	4			
## 13-1982	1.3444617	0	1	0	0	0	0	0	9			
## 13-1983	1.4332134	0	0	1	0	0	0	0	16			
## 13-1984	1.5681251	0	0	0	1	0	0	0	25			
## 13-1985	1.6998910	0	0	0	0	1	0	0	36			
## 13-1986	-0.7202626	0	0	0	0	0	1	0	49			
## 13-1987	1.6691879	0	0	0	0	0	0	1	64			
## 17-1980	1.6759624	0	0	0	0	0	0	0	16			
## 17-1981	1.5183982	1	0	0	0	0	0	0	25			
## 17-1982	1.5591905	0	1	0	0	0	0	0	36			
## 17-1983	1.7254101	0	0	1	0	0	0	0	49			
## 17-1984	1.6220223	0	0	0	1	0	0	0	64			
## 17-1985	1.6085882	0	0	0	0	1	0	0	81			
## 17-1986	1.5723854	0	0	0	0	0	1	0	100			
## 17-1987	1.8203338	0	0	0	0	0	0	1	121			
## 18-1980	1.5159627	0	0	0	0	0	0	0	16			
## 18-1981	1.7353791	1	0	0	0	0	0	0	25			
## 18-1982	1.6317437	0	1	0	0	0	0	0	36			
## 18-1983	1.9982288	0	0	1	0	0	0	0	49			

```
str(pdata)
```

[illegible]

[illegible]

[illegible]

```

## .. ..$ year: Factor w/ 8 levels "1980","1981",...: 1 2 3 4 5 6 7 8 1 2 ...
## $ d82      : 'pseries' Named int  0 0 1 0 0 0 0 0 0 0 ...
## ..- attr(*, "names")= chr  "13-1980" "13-1981" "13-1982" "13-1983" ...
## ..- attr(*, "index")=Classes 'pindex' and 'data.frame':  4360 obs. of  2 variables:
## .. ..$ nr   : Factor w/ 545 levels "13","17","18",...: 1 1 1 1 1 1 1 1 2 2 ...
## .. ..$ year: Factor w/ 8 levels "1980","1981",...: 1 2 3 4 5 6 7 8 1 2 ...
## $ d83      : 'pseries' Named int  0 0 0 1 0 0 0 0 0 0 ...
## ..- attr(*, "names")= chr  "13-1980" "13-1981" "13-1982" "13-1983" ...
## ..- attr(*, "index")=Classes 'pindex' and 'data.frame':  4360 obs. of  2 variables:
## .. ..$ nr   : Factor w/ 545 levels "13","17","18",...: 1 1 1 1 1 1 1 1 2 2 ...
## .. ..$ year: Factor w/ 8 levels "1980","1981",...: 1 2 3 4 5 6 7 8 1 2 ...
## $ d84      : 'pseries' Named int  0 0 0 0 1 0 0 0 0 0 ...
## ..- attr(*, "names")= chr  "13-1980" "13-1981" "13-1982" "13-1983" ...
## ..- attr(*, "index")=Classes 'pindex' and 'data.frame':  4360 obs. of  2 variables:
## .. ..$ nr   : Factor w/ 545 levels "13","17","18",...: 1 1 1 1 1 1 1 1 2 2 ...
## .. ..$ year: Factor w/ 8 levels "1980","1981",...: 1 2 3 4 5 6 7 8 1 2 ...
## $ d85      : 'pseries' Named int  0 0 0 0 0 1 0 0 0 0 ...
## ..- attr(*, "names")= chr  "13-1980" "13-1981" "13-1982" "13-1983" ...
## ..- attr(*, "index")=Classes 'pindex' and 'data.frame':  4360 obs. of  2 variables:
## .. ..$ nr   : Factor w/ 545 levels "13","17","18",...: 1 1 1 1 1 1 1 1 2 2 ...
## .. ..$ year: Factor w/ 8 levels "1980","1981",...: 1 2 3 4 5 6 7 8 1 2 ...
## $ d86      : 'pseries' Named int  0 0 0 0 0 0 1 0 0 0 ...
## ..- attr(*, "names")= chr  "13-1980" "13-1981" "13-1982" "13-1983" ...
## ..- attr(*, "index")=Classes 'pindex' and 'data.frame':  4360 obs. of  2 variables:
## .. ..$ nr   : Factor w/ 545 levels "13","17","18",...: 1 1 1 1 1 1 1 1 2 2 ...
## .. ..$ year: Factor w/ 8 levels "1980","1981",...: 1 2 3 4 5 6 7 8 1 2 ...
## $ d87      : 'pseries' Named int  0 0 0 0 0 0 0 1 0 0 ...
## ..- attr(*, "names")= chr  "13-1980" "13-1981" "13-1982" "13-1983" ...
## ..- attr(*, "index")=Classes 'pindex' and 'data.frame':  4360 obs. of  2 variables:
## .. ..$ nr   : Factor w/ 545 levels "13","17","18",...: 1 1 1 1 1 1 1 1 2 2 ...
## .. ..$ year: Factor w/ 8 levels "1980","1981",...: 1 2 3 4 5 6 7 8 1 2 ...
## $ expersq : 'pseries' Named int  1 4 9 16 25 36 49 64 16 25 ...
## ..- attr(*, "names")= chr  "13-1980" "13-1981" "13-1982" "13-1983" ...
## ..- attr(*, "index")=Classes 'pindex' and 'data.frame':  4360 obs. of  2 variables:
## .. ..$ nr   : Factor w/ 545 levels "13","17","18",...: 1 1 1 1 1 1 1 1 2 2 ...
## .. ..$ year: Factor w/ 8 levels "1980","1981",...: 1 2 3 4 5 6 7 8 1 2 ...
## - attr(*, "time.stamp")= chr "25 Jun 2011 23:03"
## - attr(*, "index")=Classes 'pindex' and 'data.frame':  4360 obs. of  2 variables:
## ..$ nr   : Factor w/ 545 levels "13","17","18",...: 1 1 1 1 1 1 1 1 2 2 ...
## ..$ year: Factor w/ 8 levels "1980","1981",...: 1 2 3 4 5 6 7 8 1 2 ...

```

```

head(attr(pdata, "index"),n = 20) # para extraer los ?ncices

```

```
##      nr year
## 1  13 1980
## 2  13 1981
## 3  13 1982
## 4  13 1983
## 5  13 1984
## 6  13 1985
## 7  13 1986
## 8  13 1987
## 9  17 1980
## 10 17 1981
## 11 17 1982
## 12 17 1983
## 13 17 1984
## 14 17 1985
## 15 17 1986
## 16 17 1987
## 17 18 1980
## 18 18 1981
## 19 18 1982
## 20 18 1983
```

## 2 Estadística descriptiva

```
# Descriptive statistics
summary(pdata$lwage) # proporciona informaci?n sobre porcentaje de la variabilidad que proviene de i
d y de year
```

```
## total sum of squares: 1236.53
##          id          time
## 0.53737213 0.07518366
```

```
head(pdata$lwage,n=20)
```

```
##      13-1980      13-1981      13-1982      13-1983      13-1984      13-1985
## 1.1975402 1.8530600 1.3444617 1.4332134 1.5681251 1.6998910
##      13-1986      13-1987      17-1980      17-1981      17-1982      17-1983
## -0.7202626 1.6691879 1.6759624 1.5183982 1.5591905 1.7254101
##      17-1984      17-1985      17-1986      17-1987      18-1980      18-1981
## 1.6220223 1.6085882 1.5723854 1.8203338 1.5159627 1.7353791
##      18-1982      18-1983
## 1.6317437 1.9982288
```

```
#Otra forma de mostrar los datos
head(as.matrix(pdata$lwage))
```

```
##          1980      1981      1982      1983      1984      1985      1986
## 13  1.1975402  1.853060  1.344462  1.433213  1.568125  1.699891 -0.7202626
## 17  1.6759624  1.518398  1.559191  1.725410  1.622022  1.608588  1.5723854
## 18  1.5159627  1.735379  1.631744  1.998229  2.184014  2.266662  2.0699439
## 45  1.8941150  1.471159  1.473498  1.740914  1.823214  1.908273  1.7424474
## 110 1.9487748  1.962259  1.963297  2.202516  2.134954  2.125823  1.9910166
## 120 0.2585549  1.319972  1.462085  1.320287  1.703647  1.445803  1.8063641
##          1987
## 13  1.669188
## 17  1.820334
## 18  2.873161
## 45  2.135689
## 110 2.112393
## 120 2.148120
```

```
# between (con min?scula) proporciona la media temporal de cada individuo
# y proporciona por tanto n datos (n= n?mero de individuos)
head(between(pdata$lwage),n=20) # media de cada individuo
```

```
##          13          17          18          45          110          120          126          150
## 1.255652  1.637786  2.034387  1.773664  2.055129  1.433104  1.994076  1.063356
##          162          166          189          193          209          212          218          243
## 1.472439  1.393478  1.384747  2.191872  1.838667  2.058544  2.453606  1.673951
##          259          260          309          351
## 1.695940  2.032302  2.212290  1.524323
```

```
# Between (con may?scula) proporciona tambi?n la media temporal de cada individuo
# pero ahora la salida es de tama?o n*T
head(Between(pdata$lwage),n=20) # media de cada individuo
```

```
##          13          13          13          13          13          13          13          13
## 1.255652  1.255652  1.255652  1.255652  1.255652  1.255652  1.255652  1.255652
##          17          17          17          17          17          17          17          17
## 1.637786  1.637786  1.637786  1.637786  1.637786  1.637786  1.637786  1.637786
##          18          18          18          18
## 2.034387  2.034387  2.034387  2.034387
```

```
#Within (con mayuscula) transforma cada serie rest?ndole a cada individuo su media temporal
head(Within(pdata$lwage),n=20) # dato del individuo en diferencias respecto a la media de cada indiv
iduo
```

```
##          13-1980      13-1981      13-1982      13-1983      13-1984      13-1985
## -0.05811191  0.59740793  0.08880960  0.17756128  0.31247305  0.44423889
##          13-1986      13-1987      17-1980      17-1981      17-1982      17-1983
## -1.97591466  0.41353583  0.03817607 -0.11938821 -0.07859586  0.08762373
##          17-1984      17-1985      17-1986      17-1987      18-1980      18-1981
## -0.01576410 -0.02919815 -0.06540094  0.18254746 -0.51842421 -0.29900783
##          18-1982      18-1983
## -0.40264326 -0.03615814
```

```
# diff(), lag() y lead() calcula la primera diferencia temporal,
# lag el retardo (para atr?s con k>0), y lead(el adelanto temporal con k>0)
# pero teniendo en cuenta la estructura de datos de panel
```

```
head(diff(pdata$lwage,lag=1),n=20) # media de cada individuo
```

```
##      13-1980      13-1981      13-1982      13-1983      13-1984      13-1985
##      NA      0.65551984 -0.50859833  0.08875167  0.13491178  0.13176584
##      13-1986      13-1987      17-1980      17-1981      17-1982      17-1983
## -2.42015356  2.38945049      NA -0.15756428  0.04079235  0.16621959
##      17-1984      17-1985      17-1986      17-1987      18-1980      18-1981
## -0.10338783 -0.01343405 -0.03620279  0.24794841      NA  0.21941638
##      18-1982      18-1983
## -0.10363543  0.36648512
```

```
head(lag(pdata$lwage,k=1),n=20) # media de cada individuo
```

```
##      13-1980      13-1981      13-1982      13-1983      13-1984      13-1985
##      NA      1.1975402  1.8530600  1.3444617  1.4332134  1.5681251
##      13-1986      13-1987      17-1980      17-1981      17-1982      17-1983
## 1.6998910 -0.7202626      NA  1.6759624  1.5183982  1.5591905
##      17-1984      17-1985      17-1986      17-1987      18-1980      18-1981
## 1.7254101  1.6220223  1.6085882  1.5723854      NA  1.5159627
##      18-1982      18-1983
## 1.7353791  1.6317437
```

```
head(lead(pdata$lwage,k=1),n=20) # media de cada individuo
```

```
##      13-1980      13-1981      13-1982      13-1983      13-1984      13-1985
## 1.8530600  1.3444617  1.4332134  1.5681251  1.6998910 -0.7202626
##      13-1986      13-1987      17-1980      17-1981      17-1982      17-1983
## 1.6691879      NA  1.5183982  1.5591905  1.7254101  1.6220223
##      17-1984      17-1985      17-1986      17-1987      18-1980      18-1981
## 1.6085882  1.5723854  1.8203338      NA  1.7353791  1.6317437
##      18-1982      18-1983
## 1.9982288  2.1840143
```

```
# '# Estimaci3n de los modelos de datos de PAnel
# Estimaci3n de los modelos con datos de Panel
# existen 4 m3todos en la librer?a pml
# (1) pvcmm: para modelos con todos los coeficientes variables;
# (2) plm: para modelos "between" efectos fijos o "within", "pooling" todos los coeficientes iguales,
#      "fd" first diferencias, "random" efectos aleatorios (GLS);
# (3) pgmm: m3todo de los momentos por ejemplo para modelos din3micos Arellano-Bond (no los veremos)
# (4) pggls: para ganar precisi3n estimando varianzas por cluster de individuos

# "effect"= De forma general se pueden incluir efectos individuales (fijos en el tiempo),
# efectos temporales fijos entre individuos, o ambos

# plm(formula, data, subset, weights, na.action,
#      effect = c("individual", "time", "twoways", "nested"),
#      model = c("within", "random", "ht", "between", "pooling", "fd"))
```



# 3 Pooled OLS estimator

```
# Pooled OLS estimator
# No considera datos de panel, es como poner todos los datos seguidos y considerar una constante y todas las betas iguales

modelo<- lwage~ educ + black + hisp + exper + I(exper^2) + union + married

pooling <- plm(modelo, data=pdata, model= "pooling")
summary(pooling)
```

```
## Pooling Model
##
## Call:
## plm(formula = modelo, data = pdata, model = "pooling")
##
## Balanced Panel: n = 545, T = 8, N = 4360
##
## Residuals:
##      Min.      1st Qu.      Median      3rd Qu.      Max.
## -5.268937 -0.248691  0.033205  0.296163  2.560777
##
## Coefficients:
##              Estimate Std. Error t-value Pr(>|t|)
## (Intercept) -0.03470569  0.06456900 -0.5375  0.5910
## educ         0.09938779  0.00467760 21.2476 < 2.2e-16 ***
## black       -0.14384171  0.02355950 -6.1055 1.114e-09 ***
## hisp        0.01569798  0.02081119  0.7543  0.4507
## exper       0.08917907  0.01011105  8.8200 < 2.2e-16 ***
## I(exper^2)  -0.00284866  0.00070736 -4.0272 5.742e-05 ***
## union       0.18007257  0.01712053 10.5179 < 2.2e-16 ***
## married     0.10766558  0.01569647  6.8592 7.897e-12 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Total Sum of Squares:    1236.5
## Residual Sum of Squares: 1005.8
## R-Squared:              0.18659
## Adj. R-Squared: 0.18528
## F-statistic: 142.613 on 7 and 4352 DF, p-value: < 2.22e-16
```

```
stargazer(pooling,type = "text")
```

```
##
## =====
##               Dependent variable:
##               -----
##               lwage
## -----
## educ          0.099***
##               (0.005)
##
## black         -0.144***
##               (0.024)
##
## hisp          0.016
##               (0.021)
##
## exper         0.089***
##               (0.010)
##
## I(exper2)     -0.003***
##               (0.001)
##
## union         0.180***
##               (0.017)
##
## married       0.108***
##               (0.016)
##
## Constant      -0.035
##               (0.065)
## -----
## Observations   4,360
## R2             0.187
## Adjusted R2    0.185
## F Statistic    142.613*** (df = 7; 4352)
## =====
## Note:          *p<0.1; **p<0.05; ***p<0.01
```

```
# daría el mismo resultado que estimar por lm directamente del data.frame inicial de datos
pooling2 <- lm(modelo, data=wagepan)
stargazer(pooling,pooling2,type = "text")
```

```
##
## =====
##                               Dependent variable:
##                               -----
##                               lwage
##                               panel      OLS
##                               linear
##                               (1)        (2)
## -----
## educ                0.099***        0.099***
##                    (0.005)          (0.005)
##
## black              -0.144***        -0.144***
##                    (0.024)          (0.024)
##
## hisp                0.016           0.016
##                    (0.021)          (0.021)
##
## exper              0.089***        0.089***
##                    (0.010)          (0.010)
##
## I(exper2)          -0.003***        -0.003***
##                    (0.001)          (0.001)
##
## union              0.180***        0.180***
##                    (0.017)          (0.017)
##
## married            0.108***        0.108***
##                    (0.016)          (0.016)
##
## Constant           -0.035          -0.035
##                    (0.065)          (0.065)
##
## -----
## Observations        4,360          4,360
## R2                  0.187           0.187
## Adjusted R2         0.185           0.185
## Residual Std. Error    0.481 (df = 4352)
## F Statistic (df = 7; 4352) 142.613*** 142.613***
## =====
## Note:                *p<0.1; **p<0.05; ***p<0.01
```

```
pooling3 <- plm(modelo, data=pdata, model= "pooling",effect = "twoways")
summary(pooling3)
```

```
## Pooling Model
##
## Call:
## plm(formula = modelo, data = pdata, effect = "twoways", model = "pooling")
##
## Balanced Panel: n = 545, T = 8, N = 4360
##
## Residuals:
##      Min.    1st Qu.    Median    3rd Qu.    Max.
## -5.268937 -0.248691  0.033205  0.296163  2.560777
##
## Coefficients:
##              Estimate Std. Error t-value Pr(>|t|)
## (Intercept) -0.03470569  0.06456900 -0.5375   0.5910
## educ         0.09938779  0.00467760 21.2476 < 2.2e-16 ***
## black       -0.14384171  0.02355950 -6.1055 1.114e-09 ***
## hisp        0.01569798  0.02081119  0.7543   0.4507
## exper       0.08917907  0.01011105  8.8200 < 2.2e-16 ***
## I(exper^2)  -0.00284866  0.00070736 -4.0272 5.742e-05 ***
## union       0.18007257  0.01712053 10.5179 < 2.2e-16 ***
## married     0.10766558  0.01569647  6.8592 7.897e-12 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Total Sum of Squares:    1236.5
## Residual Sum of Squares: 1005.8
## R-Squared:              0.18659
## Adj. R-Squared: 0.18528
## F-statistic: 142.613 on 7 and 4352 DF, p-value: < 2.22e-16
```

## 4 Between estimator

```
# Between estimator
# Estima el modelo pero con las medias de cada id,
# es una estimaci?n con datos de secci?n cruzada
# problema
#  $y_{bar} = \beta \cdot x_{bar} + \beta_0 + \epsilon_i + u_i$ ,  $\epsilon_i + u_i$  es no observable y es el error compuesto
# el problema del estimador Between es que est? sesgado si  $Cov(x, \epsilon_i)$  no es cero
# y si es cero, tampoco es bueno, mejor utilizar Random effects porque Between
# no recoge bien el efecto de las variables que var?an en el tiempo

between <- plm(modelo, data=pdata, model= "between")
summary(between)
```

```
## Oneway (individual) effect Between Model
##
## Call:
## plm(formula = modelo, data = pdata, model = "between")
##
## Balanced Panel: n = 545, T = 8, N = 4360
## Observations used in estimation: 545
##
## Residuals:
##      Min.      1st Qu.      Median      3rd Qu.      Max.
## -1.124205 -0.238940  0.024621  0.230350  1.737517
##
## Coefficients:
##              Estimate Std. Error t-value Pr(>|t|)
## (Intercept)  0.4923090  0.2210094  2.2275 0.0263243 *
## educ         0.0946036  0.0109043  8.6758 < 2.2e-16 ***
## black       -0.1388124  0.0488709 -2.8404 0.0046768 **
## hisp        0.0047758  0.0426925  0.1119 0.9109723
## exper       -0.0504371  0.0503326 -1.0021 0.3167578
## I(exper^2)   0.0051245  0.0032118  1.5955 0.1111867
## union       0.2706765  0.0465645  5.8129 1.054e-08 ***
## married     0.1436637  0.0411983  3.4871 0.0005282 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Total Sum of Squares:    83.06
## Residual Sum of Squares: 64.852
## R-Squared:              0.21922
## Adj. R-Squared: 0.20904
## F-statistic: 21.5386 on 7 and 537 DF, p-value: < 2.22e-16
```

```
stargazer(pooling, between,type = "text")
```

```
##
## =====
##                               Dependent variable:
##                               -----
##                               lwage
##                               (1)          (2)
## -----
## educ          0.099***          0.095***
##                (0.005)          (0.011)
##
## black         -0.144***          -0.139***
##                (0.024)          (0.049)
##
## hisp          0.016             0.005
##                (0.021)          (0.043)
##
## exper          0.089***          -0.050
##                (0.010)          (0.050)
##
## I(exper2)      -0.003***          0.005
##                (0.001)          (0.003)
##
## union          0.180***          0.271***
##                (0.017)          (0.047)
##
## married        0.108***          0.144***
##                (0.016)          (0.041)
##
## Constant      -0.035             0.492**
##                (0.065)          (0.221)
##
## -----
## Observations    4,360             545
## R2              0.187             0.219
## Adjusted R2     0.185             0.209
## F Statistic 142.613*** (df = 7; 4352) 21.539*** (df = 7; 537)
## =====
## Note:                *p<0.1; **p<0.05; ***p<0.01
```

## 5 Efectos fijos y primeras diferencias

```
# yit=beta*xit + bo + ci+uit, ci es no observable y puede considerarse
# como parte del error compuesto, por tanto la estimaci3n de beta depender?
# de si Cov(xit,ci)!=0 para alguna de las variables explicativas.

# Para estos casos Cov(xit,ci)!=0 tenemos dos estimadores consistentes

# El problema es que todas aquellas fijas en el tiempo desaparecen, no solo los efectos heterogeneos
# no
# observables, tambi3n los observables
```

### 5.1 Estimador de primeras diferencias

```
# First differences estimator
firstdiff <-plm(modelo, data=pdata, model= "fd")
summary(firstdiff)
```

```
## Oneway (individual) effect First-Difference Model
##
## Call:
## plm(formula = modelo, data = pdata, model = "fd")
##
## Balanced Panel: n = 545, T = 8, N = 4360
## Observations used in estimation: 3815
##
## Residuals:
##      Min.      1st Qu.      Median      3rd Qu.      Max.
## -4.583844 -0.146169 -0.012659  0.132687  4.836556
##
## Coefficients:
##              Estimate Std. Error t-value Pr(>|t|)
## (Intercept)  0.1157500  0.0195867  5.9096 3.729e-09 ***
## I(exper^2)   -0.0038824  0.0013863 -2.8005 0.005128 **
## union        0.0427878  0.0196575  2.1767 0.029566 *
## married      0.0381377  0.0229283  1.6633 0.096325 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Total Sum of Squares:    751.19
## Residual Sum of Squares: 748.04
## R-Squared:              0.0042035
## Adj. R-Squared: 0.0034196
## F-statistic: 5.36242 on 3 and 3811 DF, p-value: 0.0011047
```

## 5.2 Estimador de efectos Fijos

```
# Fixed effects or within estimator
fixed <- plm(modelo, data=pdata, model= "within")
summary(fixed)
```

```
## Oneway (individual) effect Within Model
##
## Call:
## plm(formula = modelo, data = pdata, model = "within")
##
## Balanced Panel: n = 545, T = 8, N = 4360
##
## Residuals:
##      Min.      1st Qu.      Median      3rd Qu.      Max.
## -4.1726214 -0.1257010  0.0092527  0.1595770  1.4701690
##
## Coefficients:
##              Estimate Std. Error t-value Pr(>|t|)
## exper        0.11684669  0.00841968 13.8778 < 2.2e-16 ***
## I(exper^2)   -0.00430089  0.00060527 -7.1057 1.422e-12 ***
## union        0.08208713  0.01929073  4.2553 2.138e-05 ***
## married      0.04530332  0.01830968  2.4743 0.01339 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Total Sum of Squares:    572.05
## Residual Sum of Squares: 470.2
## R-Squared:              0.17804
## Adj. R-Squared: 0.059852
## F-statistic: 206.375 on 4 and 3811 DF, p-value: < 2.22e-16
```

```
stargazer(pooling, between, fixed ,type = "text")
```

```
##
## =====
##                               Dependent variable:
##                               -----
##                               lwage
##                               (1)          (2)          (3)
## -----
## educ          0.099***          0.095***
##                (0.005)          (0.011)
##
## black         -0.144***          -0.139***
##                (0.024)          (0.049)
##
## hisp          0.016             0.005
##                (0.021)          (0.043)
##
## exper          0.089***          -0.050          0.117***
##                (0.010)          (0.050)          (0.008)
##
## I(exper2)      -0.003***          0.005          -0.004***
##                (0.001)          (0.003)          (0.001)
##
## union          0.180***          0.271***          0.082***
##                (0.017)          (0.047)          (0.019)
##
## married        0.108***          0.144***          0.045**
##                (0.016)          (0.041)          (0.018)
##
## Constant       -0.035             0.492**
##                (0.065)          (0.221)
##
## -----
## Observations    4,360             545             4,360
## R2              0.187             0.219             0.178
## Adjusted R2     0.185             0.209             0.060
## F Statistic 142.613*** (df = 7; 4352) 21.539*** (df = 7; 537) 206.375*** (df = 4; 3811)
## =====
## Note:                                     *p<0.1; **p<0.05; ***p<0.01
```

```
# los efectos temporales se pueden estimar sin variables dummies as?:
fixed.2 <- plm(modelo, data=pdata, model= "within", effect = "twoways")
summary(fixed.2)
```



```
## Twoways effects Within Model
##
## Call:
## plm(formula = modelo, data = pdata, effect = "twoways", model = "within")
##
## Balanced Panel: n = 545, T = 8, N = 4360
##
## Residuals:
##      Min.      1st Qu.      Median      3rd Qu.      Max.
## -4.159280 -0.125273  0.011267  0.154869  1.492088
##
## Coefficients:
##              Estimate Std. Error t-value Pr(>|t|)
## I(exper^2) -0.00518550  0.00070444 -7.3612 2.222e-13 ***
## union      0.08000186  0.01931031  4.1430 3.503e-05 ***
## married    0.04668036  0.01831044  2.5494 0.01083 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Total Sum of Squares:    479.09
## Residual Sum of Squares: 468.75
## R-Squared:              0.021568
## Adj. R-Squared:        -0.12089
## F-statistic: 27.959 on 3 and 3805 DF, p-value: < 2.22e-16
```

```
# para extraer los efectos fijos
head(summary(fixef(fixed, type="level")),10)
```

```
##      Estimate Std. Error  t-value    Pr(>|t|)
## 13  0.8292537  0.1264464  6.558146 6.179650e-11
## 17  1.0259409  0.1277434  8.031265 1.274177e-15
## 18  1.3772381  0.1274907 10.802655 8.147348e-27
## 45  1.2575037  0.1269759  9.903485 7.571453e-23
## 110 1.3623387  0.1273758 10.695431 2.517471e-26
## 120 0.9431288  0.1270172  7.425207 1.381724e-13
## 126 1.4871125  0.1267317 11.734334 2.914404e-31
## 150 0.5631199  0.1270429  4.432517 9.576104e-06
## 162 0.9722026  0.1270429  7.652553 2.480445e-14
## 166 0.8408723  0.1273605  6.602299 4.607264e-11
```

```
head(summary(fixef(fixed, type="dmean")),10) #dmean desviaciones respecto a la media global
```

```
##      Estimate Std. Error  t-value    Pr(>|t|)
## 13  -0.23562610  0.1264464 -1.8634469 6.247627e-02
## 17  -0.03893898  0.1277434 -0.3048219 7.605184e-01
## 18   0.31235826  0.1274907  2.4500472 1.432842e-02
## 45   0.19262384  0.1269759  1.5170112 1.293468e-01
## 110  0.29745881  0.1273758  2.3352859 1.958012e-02
## 120 -0.12175101  0.1270172 -0.9585397 3.378515e-01
## 126  0.42223261  0.1267317  3.3317040 8.713856e-04
## 150 -0.50175991  0.1270429 -3.9495307 7.972028e-05
## 162 -0.09267723  0.1270429 -0.7294954 4.657435e-01
## 166 -0.22400759  0.1273605 -1.7588463 7.868386e-02
```

```
# cuando tengo los dos efectos
fixef(fixed.2, effect = "time", type="level")
```

```
##      1980      1981      1982      1983      1984      1985      1986      1987
## 1.426019 1.577210 1.678989 1.780462 1.916133 2.043501 2.191515 2.351043
```

```
summary(fixef(fixed.2, effect = "time", type="dmean"))
```

```
##      Estimate Std. Error  t-value Pr(>|t|)
## 1980 -0.444591    0.018341 -24.2396 < 2.2e-16 ***
## 1981 -0.293399    0.021616 -13.5734 < 2.2e-16 ***
## 1982 -0.191620    0.026542  -7.2195 6.268e-13 ***
## 1983 -0.090147    0.033318  -2.7057 0.0068467 **
## 1984  0.045524    0.041672   1.0924 0.2747047
## 1985  0.172892    0.051544   3.3543 0.0008035 ***
## 1986  0.320906    0.063026   5.0916 3.722e-07 ***
## 1987  0.480434    0.076167   6.3076 3.158e-10 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

## 6 Estimador de Efectos Aleatorios

```
# Para los casos Cov(xit,ci)=0 entonces no hay problemas de inconsistencia
```

```
# Random effects estimator
random <- plm(modelo, data=pdata, model= "random",effect = "twoways" )
summary(random)
```

```
## Twoways effects Random Effect Model
## (Swamy-Arora's transformation)
##
## Call:
## plm(formula = modelo, data = pdata, effect = "twoways", model = "random")
##
## Balanced Panel: n = 545, T = 8, N = 4360
##
## Effects:
##               var    std.dev share
## idiosyncratic 0.1231940 0.3509900 0.539
## individual    0.1053672 0.3246031 0.461
## time          0.0001535 0.0123910 0.001
## theta: 0.6429 (id) 0.2283 (time) 0.2138 (total)
##
## Residuals:
##      Min.    1st Qu.    Median    3rd Qu.    Max.
## -4.576664 -0.144003  0.022483  0.185350  1.541723
##
## Coefficients:
##              Estimate Std. Error z-value Pr(>|z|)
## (Intercept) -0.10512206  0.11330018 -0.9278 0.3535014
## educ         0.10050264  0.00897459 11.1986 < 2.2e-16 ***
## black        -0.14385858  0.04762678 -3.0205 0.0025232 **
## hisp         0.02048311  0.04261000  0.4807 0.6307217
## exper        0.11505420  0.00887460 12.9644 < 2.2e-16 ***
## I(exper^2)   -0.00432741  0.00062404 -6.9345 4.077e-12 ***
## union        0.10670162  0.01783342  5.9832 2.187e-09 ***
## married      0.06314632  0.01676962  3.7655 0.0001662 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Total Sum of Squares:    619.18
## Residual Sum of Squares: 539.22
## R-Squared:              0.12914
## Adj. R-Squared: 0.12774
## Chisq: 645.378 on 7 DF, p-value: < 2.22e-16
```

```
stargazer(pooling2, fixed.2 , random, type = "text")
```

```
##
## =====
##                               Dependent variable:
##                               -----
##                               lwage
##                               OLS                               panel
##                               (1)                               linear
##                               (2)                               (3)
## -----
## educ                        0.099***                        0.101***
##                               (0.005)                        (0.009)
##
## black                       -0.144***                       -0.144***
##                               (0.024)                        (0.048)
##
## hisp                        0.016                           0.020
##                               (0.021)                        (0.043)
##
## exper                       0.089***                        0.115***
##                               (0.010)                        (0.009)
##
## I(exper2)                   -0.003***                       -0.005***
##                               (0.001)                        (0.001)
##
## union                       0.180***                        0.107***
##                               (0.017)                        (0.018)
##
## married                     0.108***                        0.063***
##                               (0.016)                        (0.017)
##
## Constant                    -0.035                          -0.105
##                               (0.065)                        (0.113)
##
## -----
## Observations                4,360                          4,360
## R2                          0.187                          0.129
## Adjusted R2                 0.185                          -0.121
## Residual Std. Error         0.481 (df = 4352)
## F Statistic                  142.613*** (df = 7; 4352) 27.959*** (df = 3; 3805) 645.378***
## =====
## Note:                        *p<0.1; **p<0.05; ***p<0.01
```

## 7 Diagnosis del modelo

```
#DIAGNOSIS

#cuando tengo grados de libertad suficiente es posible contrastar si todas las pendientes son iguales
# o no
# en est caso no lo podemos hacer
#mvariable<-pvcmm(modelo, data=pdata, model= "within")
#fixed <- plm(modelo, data=pdata, model= "within")
# pooltest(fixed,mvariable)

# LM test for fixed effects versus OLS
plmtest(pooling)
```

```
##
## Lagrange Multiplier Test - (Honda) for balanced panels
##
## data: modelo
## normal = 56.716, p-value < 2.2e-16
## alternative hypothesis: significant effects
```

```
head(summary(fixef(fixed, type="level")),10)
```

```
##      Estimate Std. Error  t-value    Pr(>|t|)
## 13  0.8292537  0.1264464  6.558146 6.179650e-11
## 17  1.0259409  0.1277434  8.031265 1.274177e-15
## 18  1.3772381  0.1274907 10.802655 8.147348e-27
## 45  1.2575037  0.1269759  9.903485 7.571453e-23
## 110 1.3623387  0.1273758 10.695431 2.517471e-26
## 120 0.9431288  0.1270172  7.425207 1.381724e-13
## 126 1.4871125  0.1267317 11.734334 2.914404e-31
## 150 0.5631199  0.1270429  4.432517 9.576104e-06
## 162 0.9722026  0.1270429  7.652553 2.480445e-14
## 166 0.8408723  0.1273605  6.602299 4.607264e-11
```

```
pFtest(fixed.2, pooling2)
```

```
##
## F test for twoways effects
##
## data: modelo
## F = 7.9697, df1 = 547, df2 = 3805, p-value < 2.2e-16
## alternative hypothesis: significant effects
```

```
# Hausman test for fixed versus random effects model
phtest(random, fixed.2)
```

```
##
## Hausman Test
##
## data: modelo
## chisq = 23.948, df = 3, p-value = 2.561e-05
## alternative hypothesis: one model is inconsistent
```

## 8 Corrección por hetrocedasticidad

```
# ¿qué pasa con la heteroscedasticidad?: estimación robusta de las VAR(betas)
fixed <- plm(modelo, data=pdata, model= "within", effect = "twoways")
summary(fixed)
```

```
## Twoways effects Within Model
##
## Call:
## plm(formula = modelo, data = pdata, effect = "twoways", model = "within")
##
## Balanced Panel: n = 545, T = 8, N = 4360
##
## Residuals:
##      Min.    1st Qu.      Median    3rd Qu.      Max.
## -4.159280 -0.125273  0.011267  0.154869  1.492088
##
## Coefficients:
##              Estimate Std. Error t-value Pr(>|t|)
## I(exper^2) -0.00518550  0.00070444 -7.3612 2.222e-13 ***
## union      0.08000186  0.01931031  4.1430 3.503e-05 ***
## married    0.04668036  0.01831044  2.5494  0.01083 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Total Sum of Squares:    479.09
## Residual Sum of Squares: 468.75
## R-Squared:      0.021568
## Adj. R-Squared: -0.12089
## F-statistic: 27.959 on 3 and 3805 DF, p-value: < 2.22e-16
```

```
suppressMessages(library(lmtest))
suppressMessages(library(sandwich))

fixed.ro<-coeftest(fixed, vcov = vcovHC(fixed, type = "HC1",cluster = "group",adjust = T))
# Da el mismo resultado que fixed.ro<-coeftest(fixed, vcov = vcovHC(fixed, type = "HC1"))
random.ro<-coeftest(random, vcov = vcovHC(fixed, type = "HC1",cluster = "group",adjust = T))

# para sacar el mismo resultado que stata hay que ajustar los grados de libertad
numid<-length(unique(pdata$id))
mT<-length(pdata$id)
dfa<-(numid/(numid-1))*(mT-1)/fixed$df.residual

fixed.rostata<-coeftest(fixed, vcov = dfa*vcovHC(fixed, type = "HC1",cluster = "group",adjust = T))
random.rostata<-coeftest(random, vcov = vcovHC(random, type = "HC1",cluster = "group",adjust = T))

stargazer(fixed, fixed.ro , random, random.ro, type = "text")
```

```
##
## =====
##                               Dependent variable:
##                               -----
##               lwage           lwage
##               panel           panel
##               linear           linear
##               (1)             (2)   (3)             (4)
## -----
## educ                               0.101***
##                               (0.009)
##
## black                             -0.144***
##                               (0.048)
##
## hisp                               0.020
##                               (0.043)
##
## exper                             0.115***
##                               (0.009)
##
## I(exper2)        -0.005***      -0.005***  -0.004***  -0.004***
##                (0.001)          (0.001)   (0.001)   (0.001)
##
## union            0.080***        0.080***   0.107***   0.107***
##                (0.019)          (0.023)   (0.018)   (0.023)
##
## married          0.047**         0.047**   0.063***   0.063***
##                (0.018)          (0.021)   (0.017)   (0.021)
##
## Constant                               -0.105
##                               (0.113)
##
## -----
## Observations           4,360                4,360
## R2                     0.022                0.129
## Adjusted R2            -0.121                0.128
## F Statistic  27.959*** (df = 3; 3805)        645.378***
## =====
## Note:                               *p<0.1; **p<0.05; ***p<0.01
```

#para sacar los datos con la nueva estimaci?n utilizamos la libreria lmtest

```
fixed.ro2 <- pgglsl(modelo, data=pdata, model= "within",effect = "individual")
summary(fixed.ro2)
```

```
## Oneway (individual) effect Within FGLS model
##
## Call:
## pgls(formula = modelo, data = pdata, effect = "individual",
##       model = "within")
##
## Balanced Panel: n = 545, T = 8, N = 4360
##
## Residuals:
##      Min.    1st Qu.    Median      Mean   3rd Qu.      Max.
## -4.16917 -0.12458  0.01038  0.00000  0.15973  1.47184
##
## Coefficients:
##              Estimate Std. Error z-value Pr(>|z|)
## exper          0.10926814  0.01070210 10.2100 < 2.2e-16 ***
## I(exper^2) -0.00362175  0.00073614 -4.9199 8.659e-07 ***
## union          0.06698336  0.01859042  3.6031 0.0003144 ***
## married        0.04976684  0.01886630  2.6379 0.0083429 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## Total Sum of Squares: 1236.5
## Residual Sum of Squares: 470.49
## Multiple R-squared: 0.6195
```

## 9 Estimador CRE

```
# Primera alternativa es obtener los efectos fijos y
# despues hacer una segunda regresión de estos efectos fijos
# sobre los efectos fijos observables

efijos<-data.frame(beduc=between(pdata$educ), bblack=between(pdata$black),
                  bhispanic=between(pdata$hispanic), bexper=between(pdata$exper),
                  ef=as.numeric(fixef(fixed, type="level")))
efijos.lm<-lm(ef~beduc +bblack+ bhispanic + bexper, data=efijos)
summary(efijos.lm)
```

```
##
## Call:
## lm(formula = ef ~ beduc + bblack + bhispanic + bexper, data = efijos)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.17936 -0.24427  0.00246  0.24271  1.72388
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  0.05344    0.18205   0.294  0.76922
## beduc        0.09236    0.01087   8.497 < 2e-16 ***
## bblack       -0.13965    0.04853  -2.878 0.00416 **
## bhispanic     0.02335    0.04352   0.537 0.59176
## bexper        0.11404    0.01135  10.048 < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.3561 on 540 degrees of freedom
## Multiple R-squared:  0.1799, Adjusted R-squared:  0.1738
## F-statistic: 29.61 on 4 and 540 DF, p-value: < 2.2e-16
```



## 9.1 Correlated Random Effects

```
# Segunda alternativa Correlated Random Effects
```

```
stargazer(pooling, fixed, random,type = "text")
```

```
##
## =====
##                               Dependent variable:
##                               -----
##                               lwage
##                               (1)          (2)          (3)
## -----
## educ                0.099***          0.101***
##                   (0.005)          (0.009)
##
## black              -0.144***          -0.144***
##                   (0.024)          (0.048)
##
## hisp                0.016
##                   (0.021)
##
## exper              0.089***          0.115***
##                   (0.010)          (0.009)
##
## I(exper2)          -0.003***          -0.005***          -0.004***
##                   (0.001)          (0.001)          (0.001)
##
## union              0.180***          0.080***          0.107***
##                   (0.017)          (0.019)          (0.018)
##
## married            0.108***          0.047**          0.063***
##                   (0.016)          (0.018)          (0.017)
##
## Constant           -0.035
##                   (0.065)
##
## -----
## Observations        4,360          4,360          4,360
## R2                  0.187          0.022          0.129
## Adjusted R2         0.185          -0.121          0.128
## F Statistic 142.613*** (df = 7; 4352) 27.959*** (df = 3; 3805) 645.378***
## =====
## Note:                *p<0.1; **p<0.05; ***p<0.01
```

```

#Como ni edci ni black ni hisp ni exper varían salen del modelo de efectos fijos
# Se puede obtener la estimación de Efectos fijos de estas variables fijas que sí son observables utilizando CRE

#Se hace la media de cada grupo de las variables que varían en i,t
# (ni las variables fijas ni los efectos temporales se incluyen)
#con Between mayúsculas
pdata$exper2_bar<-Between(pdata$exper^2)
pdata$union_bar<-Between(pdata$union)
pdata$married_bar<-Between(pdata$married)

# Y ahora se incluye estas medias en la regresión pero con efectos aleatorios: El resultado es el de efectos fijos

modelo_CRE<- lwage~ educ + black + hisp + exper + I(exper^2) + union + married + exper2_bar + union_bar + married_bar

CRE <- plm(modelo_CRE, data=pdata, model= "random", effect = "twoways")
summary(CRE)

```

```
## Twoways effects Random Effect Model
##   (Swamy-Arora's transformation)
##
## Call:
## plm(formula = modelo_CRE, data = pdata, effect = "twoways", model = "random")
##
## Balanced Panel: n = 545, T = 8, N = 4360
##
## Effects:
##               var    std.dev share
## idiosyncratic 0.1231940 0.3509900 0.539
## individual    0.1053672 0.3246031 0.461
## time          0.0001535 0.0123910 0.001
## theta: 0.6429 (id) 0.2283 (time) 0.2138 (total)
##
## Residuals:
##      Min.    1st Qu.    Median    3rd Qu.    Max.
## -4.555284 -0.143467  0.022897  0.185928  1.548394
##
## Coefficients:
##              Estimate Std. Error z-value Pr(>|z|)
## (Intercept) -0.03263828  0.15590896 -0.2093 0.8341813
## educ         0.09246981  0.01090020  8.4833 < 2.2e-16 ***
## black        -0.14766902  0.04886442 -3.0220 0.0025110 **
## hisp         0.01486774  0.04264269  0.3487 0.7273457
## exper        0.11529906  0.00890661 12.9453 < 2.2e-16 ***
## I(exper^2)   -0.00424074  0.00063164 -6.7139 1.895e-11 ***
## union        0.08140843  0.01931527  4.2147 2.501e-05 ***
## married      0.04814667  0.01831141  2.6293 0.0085554 **
## exper2_bar   -0.00093988  0.00075875 -1.2387 0.2154485
## union_bar    0.16425762  0.04990738  3.2912 0.0009974 ***
## married_bar  0.08531930  0.04498276  1.8967 0.0578661 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Total Sum of Squares:    619.18
## Residual Sum of Squares: 537.2
## R-Squared:    0.13239
## Adj. R-Squared: 0.1304
## Chisq: 663.632 on 10 DF, p-value: < 2.22e-16
```

```
CRE.ro<-coefstest(CRE, vcov = vcovHC(CRE, type = "HC1",cluster = "group",adjust = T))
```

```
stargazer(fixed, random, CRE, type = "text")
```

##	=====		
##	Dependent variable:		
##	-----		
##	lwage		
##	(1)	(2)	(3)
##	-----		
## educ		0.101***	0.092***
##		(0.009)	(0.011)
##			
## black		-0.144***	-0.148***
##		(0.048)	(0.049)
##			
## hisp		0.020	0.015
##		(0.043)	(0.043)
##			
## exper		0.115***	0.115***
##		(0.009)	(0.009)
##			
## I(exper2)	-0.005***	-0.004***	-0.004***
##	(0.001)	(0.001)	(0.001)
##			
## union	0.080***	0.107***	0.081***
##	(0.019)	(0.018)	(0.019)
##			
## married	0.047**	0.063***	0.048***
##	(0.018)	(0.017)	(0.018)
##			
## exper2_bar			-0.001
##			(0.001)
##			
## union_bar			0.164***
##			(0.050)
##			
## married_bar			0.085*
##			(0.045)
##			
## Constant		-0.105	-0.033
##		(0.113)	(0.156)
##			
##	-----		
## Observations	4,360	4,360	4,360
## R2	0.022	0.129	0.132
## Adjusted R2	-0.121	0.128	0.130
## F Statistic	27.959*** (df = 3; 3805)	645.378***	663.632***
##	=====		
## Note:	*p<0.1; **p<0.05; ***p<0.01		



Proyecto de innovación  
“Innova Docencia” de la UCM  
Convocatoria 2018-2019  
Proyecto número 275

**R-adaptación de la asignatura de métodos econométricos en economía y finanzas del  
Grado de Estadística Aplicada**

Facultad de Estudios Estadísticos  
Departamento de Economía Aplicada, Pública y Política

**Título**

***Práctica 18: Regresión cuantílica para la discriminación salarial y  
distribución de salarios***

**Correspondiente al**

**Capítulo 9 MODELOS DE REGRESIÓN CUANTÍLICA**

- Regresión en media y regresión por cuantiles
- Estimación de la regresión intercuantílica
- Representación gráfica de la solución

***Práctica 18: Regresión cuantílica para la discriminación salarial y  
distribución de salarios***

# Regresión Cuantílica

## Asignatura Métodos Econométricos en Economía y Finanzas

Profesores  
Loren Escot, Alicia Pérez-Alonso y Julio E. Sandubete

Dpto Economía Aplicada, Pública y Política  
Facultad Estudios Estadísticos - Universidad Complutense de Madrid

Curso académico 2018/2019

### Abstract

En esta práctica se presenta la regresión cuantílica. A diferencia de la regresión por MCO que estima el efecto en media, la regresión cuantílica busca estimar los efectos de las variables explicativas sobre los diferentes cuantiles de la distribución de la variable Dependiente. Se presentará un ejemplo para estimar una ecuación de salarios por cuantiles.

- 1 Objetivo
- 2 Carga de librerías necesarias
- 3 Carga de los datos de salarios
- 4 Estimación de la ecuación de salarios por MCO
- 5 Gráfico de los efectos sobre los cuantiles
- 6 Regresión Cuantílica
  - 6.1 Cuantil 12.5
  - 6.2 Cuantil 25
  - 6.3 Cuantil 5
  - 6.4 Cuantil 75
  - 6.5 Cuantil 90
- 7 Diagnósis sobre los efectos sobre los diferentes cuantiles

```
#=====#  
  
rm(list=ls()) # Elimina todos los objetos que tengamos en la memoria  
#####33  
  
# Regresión Cuantílica
```

## 1 Objetivo

A diferencia de la regresión por MCO que estima el efecto en media, la regresión cuantílica busca estimar los efectos de las variables explicativas sobre los diferentes cuantiles de la distribución de la variable Dependiente. Este método de regresión es especialmente interesante cuando la distribución de la variable dependiente NO es simétrica

Ilustraremos la práctica con una regresión cuantílica de los salarios.

## 2 Carga de librerías necesarias

```
#
# Regresión Cuantílica

suppressMessages(library(dplyr))
suppressMessages(library(stargazer))

# la librería para estimar la regresió cuantílica
suppressMessages(library(quantreg))
```

### 3 Carga de los datos de salarios

```
Varones<-read.csv(file="salarios_PracticaHombres.csv",sep = ";",stringsAsFactors = FALSE)
Mujeres<-read.csv(file="salarios_PracticaMujeres.csv",sep = ";",stringsAsFactors = FALSE)

summary(Varones$SALARIO)
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##      48    1150    1430    1645    1983    8667
```

```
summary(Mujeres$SALARIO)
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##      70.0   865.3  1166.7  1333.6  1680.0  6857.5
```

```
meanV<-mean(Varones$SALARIO)
print(meanV)
```

```
## [1] 1645.131
```

```
medianV<-median(Varones$SALARIO)
print(medianV)
```

```
## [1] 1429.58
```

```
meanM<-mean(Mujeres$SALARIO)
print(meanM)
```

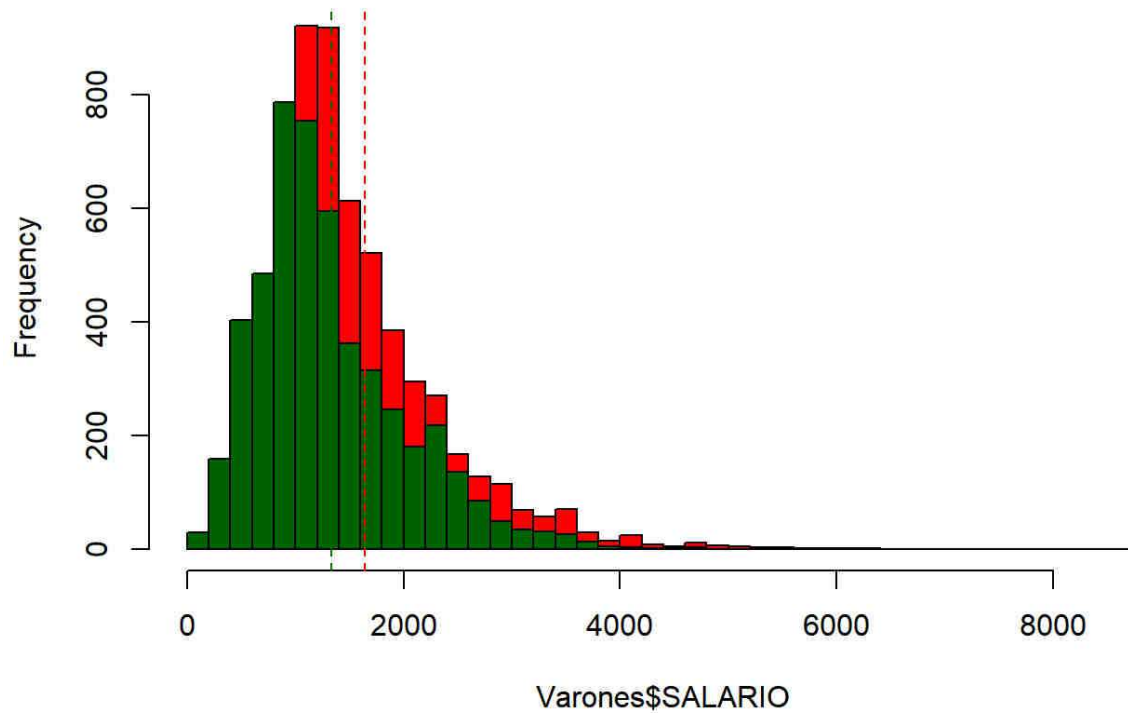
```
## [1] 1333.577
```

```
medianM<-median(Mujeres$SALARIO)
print(medianM)
```

```
## [1] 1166.66
```

```
hist(Varones$SALARIO, col="red",nclass = 40)
hist(Mujeres$SALARIO, col="darkgreen",nclass = 40, add=TRUE)
abline(v=meanV, col="red", lty=2)
abline(v=meanM, col="darkgreen", lty=2)
```

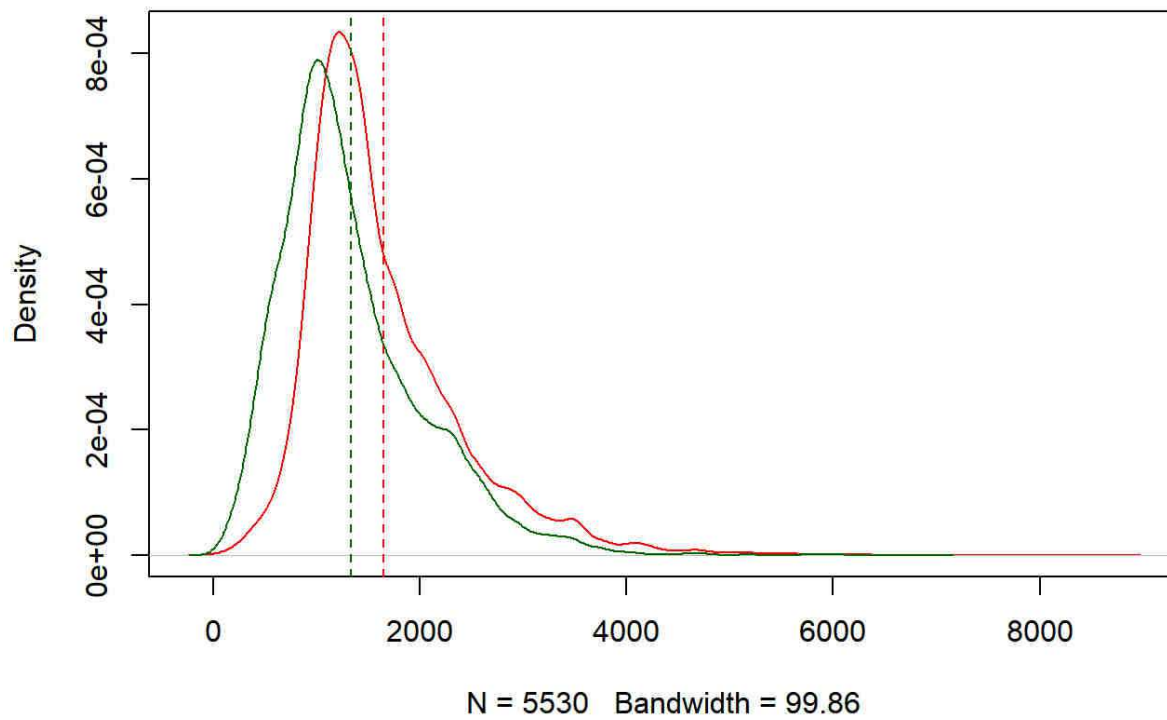
## Histogram of Varones\$SALARIO



```
plot(density(Varones$SALARIO), col="red")
lines(density(Mujeres$SALARIO), col="darkgreen")
abline(v=meanV, col="red", lty=2)
abline(v=meanM, col="darkgreen", lty=2)
```

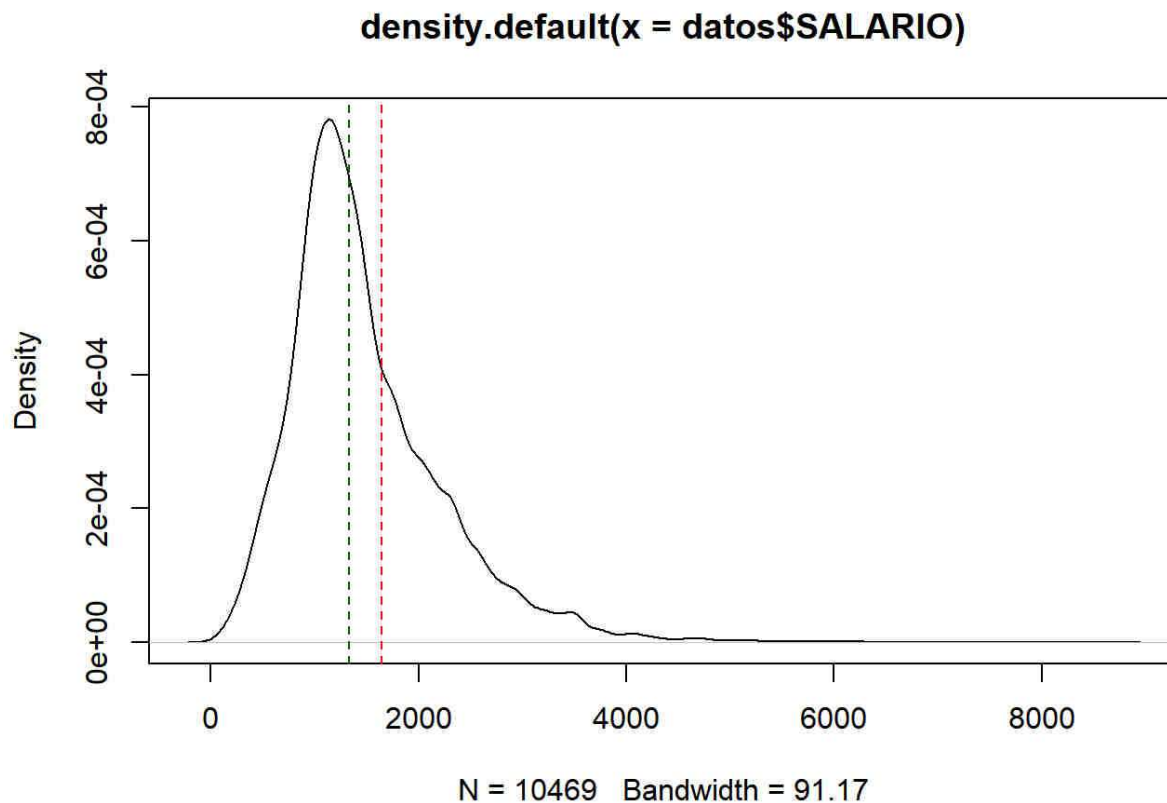


### density.default(x = Varones\$SALARIO)



```
# Unimos los datos de Varones y mujeres
datos<-rbind.data.frame(Varones, Mujeres)
```

```
plot(density(datos$SALARIO))
abline(v=meanV, col="red", lty=2)
abline(v=meanM, col="darkgreen", lty=2)
```



```
# HORAS  
datos$HORASMENSALES<-datos$HORASSEMANA*52/12
```

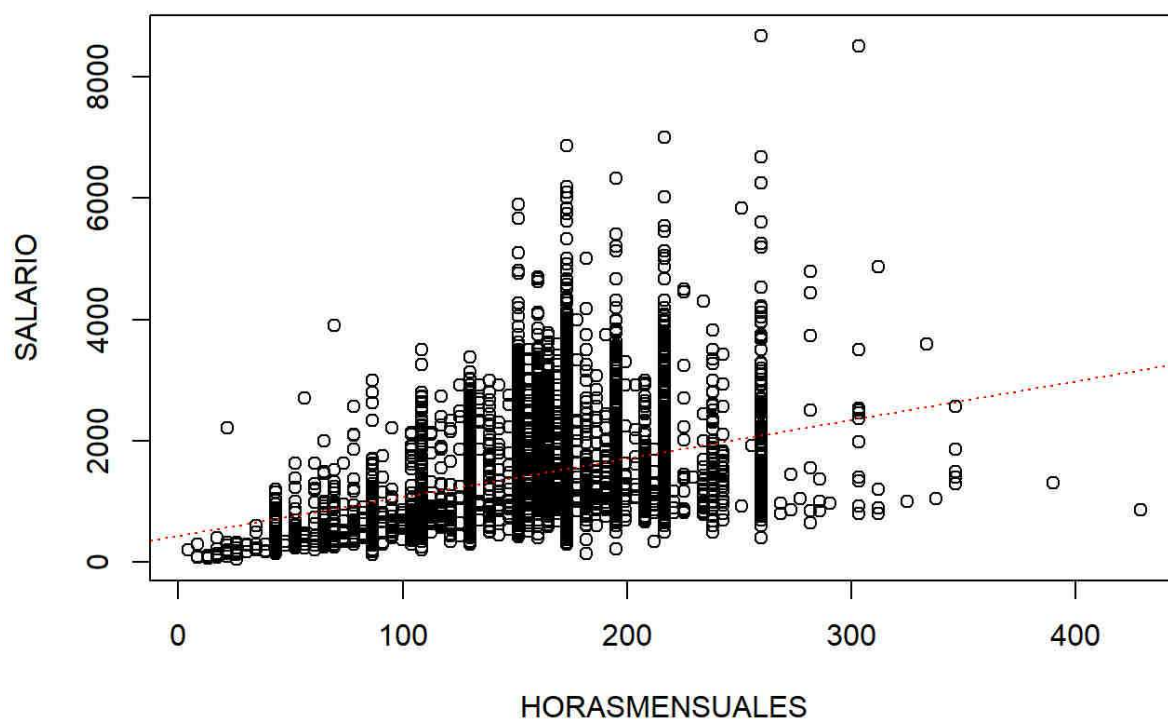
## 4 Estimación de la ecuación de salarios por MCO

```
# MODELO GENERAL  
modelo<-SALARIO ~ MUJER + ESUPERIORES + ESECUNDARIOS +  
  EXPERIENCIA + I(EXPERIENCIA^2) + ENFERMEDAD+ HORASMENSALES+ SUPERVISOR+  
  TEMPORAL+ EXTRANJERO + CASADO+ CASADO*MUJER+ Numerohijos12 + Numerohijos12*MUJER  
  
# MODELO LINEAL MCO  
model.lm<-lm(modelo, data=datos)  
summary(model.lm)
```

```
##
## Call:
## lm(formula = modelo, data = datos)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1761.2   -345.8   -61.9    266.6   5896.8
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    -93.7805     37.9143  -2.473   0.01340 *
## MUJER          -112.5698     19.2883  -5.836 5.50e-09 ***
## ESUPERIORES     823.9837     19.7129  41.799 < 2e-16 ***
## ESECUNDARIOS    226.7586     18.7354  12.103 < 2e-16 ***
## EXPERIENCIA     26.2846      1.8828  13.961 < 2e-16 ***
## I(EXPERIENCIA^2) -0.2488      0.0415  -5.995 2.11e-09 ***
## ENFERMEDAD     -40.1165     14.4068  -2.785   0.00537 **
## HORASMENSUALES    4.3248      0.1607  26.917 < 2e-16 ***
## SUPERVISOR      322.1570     13.8267  23.300 < 2e-16 ***
## TEMPORAL       -132.4653     14.6302  -9.054 < 2e-16 ***
## EXTRANJERO     -180.3635     19.9622  -9.035 < 2e-16 ***
## CASADO          143.9959     19.6942   7.312 2.83e-13 ***
## Numerohijos12    30.7401     11.2810   2.725 0.00644 **
## MUJER:CASADO    -92.8972     25.8048  -3.600 0.00032 ***
## MUJER:Numerohijos12 -20.8803     15.7812  -1.323 0.18583
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 563.1 on 10454 degrees of freedom
## Multiple R-squared:  0.4493, Adjusted R-squared:  0.4486
## F-statistic: 609.2 on 14 and 10454 DF, p-value: < 2.2e-16
```

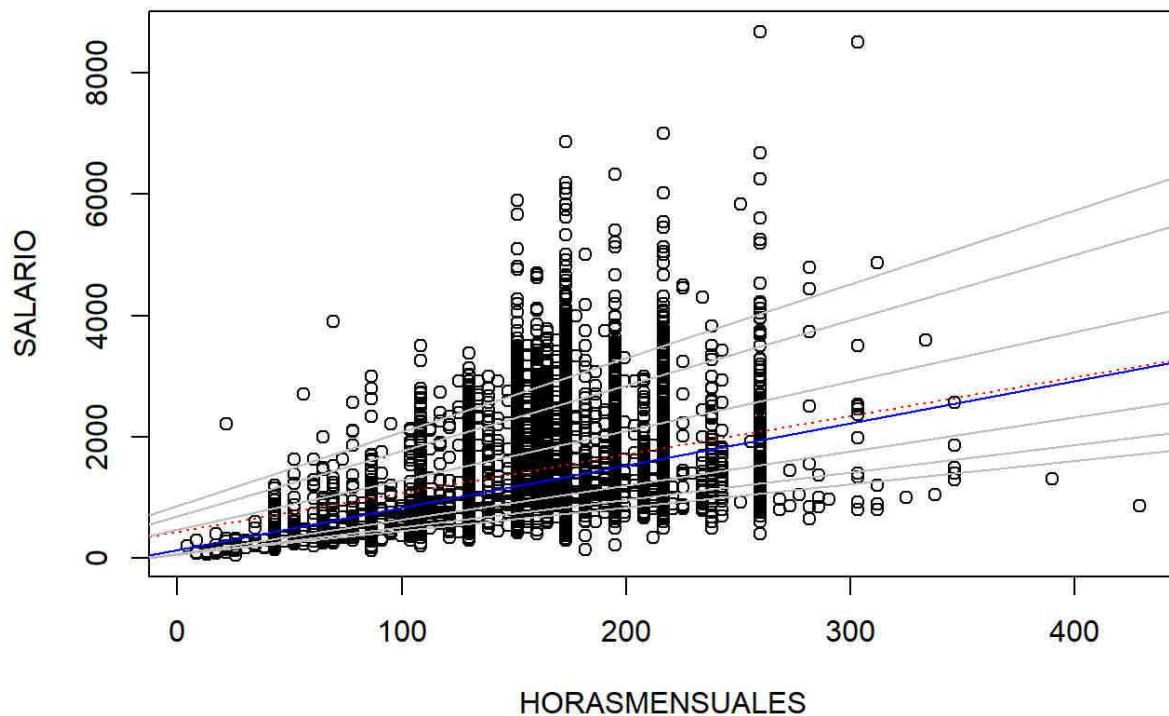
```
#Efecto sobre la media
```

```
plot(SALARIO~HORASMENSUALES,data=datos,xlab="HORASMENSUALES", ylab="SALARIO")
abline(lm(SALARIO~HORASMENSUALES, data=datos),lty = 3,col="red")
```



## 5 Gráfico de los efectos sobre los cuantiles

```
#####dibujo el efecto sobre los diferentes cuantiles
plot(SALARIO~HORASMENSALES,data=datos,xlab="HORASMENSALES", ylab="SALARIO")
taus <- c(.05,.1,.25,.75,.90,.95)
abline(rq(SALARIO~HORASMENSALES,tau=.5, data=datos),col="blue")
abline(lm(SALARIO~HORASMENSALES, data=datos),lty = 3,col="red")
for( i in 1:length(taus)){
  abline(rq(SALARIO~HORASMENSALES,tau=taus[i], data=datos),col="gray")
}
```



```
# Las pendientes son diferentes, esto es el efecto de una hora
# más de trabajo no es el mismo en el primer decil de salarios
# que en el decil_90 (una hora adicional entre los que ganan menos
# se remunera mucho menos que una hora adicional entre los que más
# ganan)
```

## 6 Regresión Cuantílica

```
# Quantile regression
```

### 6.1 Cuantil 12.5

```
quantreg125 <- rq(modelo, data=datos, tau=0.125)
```

```
## Warning in rq.fit.br(x, y, tau = tau, ...): Solution may be nonunique
```

```
summary(quantreg125, se="boot")
```

```
##
## Call: rq(formula = modelo, tau = 0.125, data = datos)
##
## tau: [1] 0.125
##
## Coefficients:
```

	Value	Std. Error	t value	Pr(> t )
(Intercept)	-32.99697	34.85564	-0.94668	0.34383
MUJER	-82.51742	16.77733	-4.91839	0.00000
ESUPERIORES	370.51257	17.62030	21.02760	0.00000
ESECUNDARIOS	114.32653	15.13961	7.55149	0.00000
EXPERIENCIA	13.31388	1.73193	7.68729	0.00000
I(EXPERIENCIA^2)	-0.16343	0.03937	-4.15077	0.00003
ENFERMEDAD	-32.07401	10.49376	-3.05648	0.00225
HORASMENSUALES	3.64530	0.13782	26.44941	0.00000
SUPERVISOR	152.10405	15.10115	10.07235	0.00000
TEMPORAL	-111.96656	11.17510	-10.01929	0.00000
EXTRANJERO	-112.35608	17.73583	-6.33498	0.00000
CASADO	125.45666	17.51389	7.16327	0.00000
Numerohijos12	-4.44839	11.09676	-0.40087	0.68852
MUJER:CASADO	-97.50987	22.52814	-4.32836	0.00002
MUJER:Numerohijos12	4.61648	13.37854	0.34507	0.73005

## 6.2 Cuantil 25

```
quantreg25 <- rq(modelo, data=datos, tau=0.25)
summary(quantreg25, se="boot")
```

```
##
## Call: rq(formula = modelo, tau = 0.25, data = datos)
##
## tau: [1] 0.25
##
## Coefficients:
```

	Value	Std. Error	t value	Pr(> t )
(Intercept)	-18.47874	30.05768	-0.61478	0.53872
MUJER	-101.76364	15.64982	-6.50254	0.00000
ESUPERIORES	459.42662	19.46362	23.60437	0.00000
ESECUNDARIOS	129.64274	15.41597	8.40964	0.00000
EXPERIENCIA	14.14300	1.49499	9.46027	0.00000
I(EXPERIENCIA^2)	-0.14974	0.03421	-4.37735	0.00001
ENFERMEDAD	-29.92180	13.71171	-2.18221	0.02912
HORASMENSUALES	4.04342	0.12983	31.14303	0.00000
SUPERVISOR	201.55306	13.70680	14.70461	0.00000
TEMPORAL	-104.54096	10.47089	-9.98396	0.00000
EXTRANJERO	-136.30311	16.22332	-8.40168	0.00000
CASADO	130.20788	17.50871	7.43675	0.00000
Numerohijos12	2.98505	10.35334	0.28832	0.77311
MUJER:CASADO	-95.47585	21.72742	-4.39426	0.00001
MUJER:Numerohijos12	-0.70042	13.98824	-0.05007	0.96007

## 6.3 Cuantil 5

```
quantreg50 <- rq(modelo, data=datos, tau=0.5)
```

```
## Warning in rq.fit.br(x, y, tau = tau, ...): Solution may be nonunique
```

```
summary(quantreg50, se="boot" )
```

```
##
## Call: rq(formula = modelo, tau = 0.5, data = datos)
##
## tau: [1] 0.5
##
## Coefficients:
##              Value      Std. Error t value    Pr(>|t|)
## (Intercept)      1.79305      31.74961    0.05647    0.95496
## MUJER           -101.40114     17.01715   -5.95876    0.00000
## ESUPERIORES      695.45174     19.72503   35.25732    0.00000
## ESECUNDARIOS     173.89174     14.10628   12.32726    0.00000
## EXPERIENCIA      17.98149      1.76841   10.16815    0.00000
## I(EXPERIENCIA^2)  -0.13639      0.04075   -3.34745    0.00082
## ENFERMEDAD       -24.59792     14.28713   -1.72168    0.08516
## HORASMENSUALES     4.33813      0.15708   27.61666    0.00000
## SUPERVISOR       270.79594     17.06780   15.86590    0.00000
## TEMPORAL        -113.80147      9.84147  -11.56346    0.00000
## EXTRANJERO       -153.90335     17.04738   -9.02798    0.00000
## CASADO           144.36844     14.66247    9.84612    0.00000
## Numerohijos12     25.98475     10.90203    2.38348    0.01717
## MUJER:CASADO     -103.95316     21.64636   -4.80234    0.00000
## MUJER:Numerohijos12 -14.24130     15.16793   -0.93891    0.34780
```

## 6.4 Cuantil 75

```
quantreg75 <- rq(modelo, data=datos, tau=0.75)
summary(quantreg75, se="boot")
```

```
##
## Call: rq(formula = modelo, tau = 0.75, data = datos)
##
## tau: [1] 0.75
##
## Coefficients:
##              Value      Std. Error t value    Pr(>|t|)
## (Intercept)     51.70152     43.58384    1.18625    0.23555
## MUJER           -121.50667     20.28755   -5.98922    0.00000
## ESUPERIORES      993.02754     23.92038   41.51386    0.00000
## ESECUNDARIOS     216.72360     17.31183   12.51882    0.00000
## EXPERIENCIA      23.38233      2.30156   10.15932    0.00000
## I(EXPERIENCIA^2)  -0.17052      0.05554   -3.07006    0.00215
## ENFERMEDAD       -32.02498     16.03519   -1.99717    0.04583
## HORASMENSUALES     4.85386      0.22287   21.77874    0.00000
## SUPERVISOR       361.02906     24.96578   14.46095    0.00000
## TEMPORAL        -116.24975     14.03556   -8.28251    0.00000
## EXTRANJERO       -161.59159     21.80630   -7.41032    0.00000
## CASADO           135.97384     28.00137    4.85597    0.00000
## Numerohijos12     36.46182     19.01592    1.91744    0.05521
## MUJER:CASADO     -96.99064     29.81099   -3.25352    0.00114
## MUJER:Numerohijos12 -19.08040     21.28456   -0.89644    0.37004
```

## 6.5 Cuantil 90

```
quantreg90 <- rq(modelo, data=datos, tau=0.90)
summary(quantreg90, se="boot")
```

```
##
## Call: rq(formula = modelo, tau = 0.9, data = datos)
##
## tau: [1] 0.9
##
## Coefficients:
##
```

	Value	Std. Error	t value	Pr(> t )
## (Intercept)	144.80414	65.71645	2.20347	0.02758
## MUJER	-186.25704	31.16069	-5.97731	0.00000
## ESUPERIORES	1190.39589	33.04646	36.02189	0.00000
## ESECUNDARIOS	282.09073	20.40910	13.82181	0.00000
## EXPERIENCIA	21.98616	3.35404	6.55512	0.00000
## I(EXPERIENCIA^2)	-0.02652	0.08456	-0.31364	0.75380
## ENFERMEDAD	-58.59154	24.42986	-2.39836	0.01649
## HORASMENSUALES	5.41206	0.32597	16.60280	0.00000
## SUPERVISOR	538.51031	39.28721	13.70701	0.00000
## TEMPORAL	-101.47458	20.79439	-4.87990	0.00000
## EXTRANJERO	-153.36720	26.92322	-5.69647	0.00000
## CASADO	148.07656	42.09448	3.51772	0.00044
## Numerohijos12	78.91168	27.19215	2.90200	0.00372
## MUJER:CASADO	-84.69192	46.15770	-1.83484	0.06656
## MUJER:Numerohijos12	-38.78823	35.12101	-1.10442	0.26944

```
##
```

```
stargazer(quantreg125,quantreg25, quantreg50,quantreg75,quantreg90, type = "text")
```



```

##
## =====
##                               Dependent variable:
##                               -----
##                               SALARIO
##                               (1)      (2)      (3)      (4)      (5)
## -----
## MUJER                -82.517***   -101.764*** -101.401*** -121.507*** -186.257***
##                      (13.780)    (12.873)    (14.563)    (22.470)    (32.891)
##
## ESUPERIORES          370.513***   459.427***  695.452***  993.028***  1,190.396***
##                      (17.268)    (14.348)    (16.982)    (23.709)    (30.123)
##
## ESECUNDARIOS          114.327***   129.643***  173.892***  216.724***  282.091***
##                      (15.213)    (10.667)    (11.124)    (19.069)    (21.769)
##
## EXPERIENCIA           13.314***    14.143***   17.981***   23.382***   21.986***
##                      (1.340)     (1.391)     (1.615)     (2.297)     (2.971)
##
## I(EXPERIENCIA2)       -0.163***   -0.150***   -0.136***   -0.171***    -0.027
##                      (0.031)     (0.033)     (0.040)     (0.055)     (0.067)
##
## ENFERMEDAD            -32.074***   -29.922**   -24.598*    -32.025*    -58.592**
##                      (9.816)     (12.402)    (13.575)    (17.336)    (23.513)
##
## HORASMENSUALES         3.645***     4.043***    4.338***    4.854***    5.412***
##                      (0.097)     (0.108)     (0.122)     (0.164)     (0.220)
##
## SUPERVISOR            152.104***   201.553***  270.796***  361.029***  538.510***
##                      (14.922)    (13.386)    (16.585)    (22.528)    (34.677)
##
## TEMPORAL              -111.967***  -104.541*** -113.801*** -116.250*** -101.475***
##                      (10.486)    (9.428)     (10.430)    (16.020)    (21.426)
##
## EXTRANJERO            -112.356***  -136.303*** -153.903*** -161.592*** -153.367***
##                      (13.990)    (14.921)    (15.445)    (22.400)    (25.395)
##
## CASADO                125.457***   130.208***  144.368***  135.974***  148.077***
##                      (14.867)    (15.852)    (17.029)    (28.010)    (38.864)
##
## Numerohijos12         -4.448        2.985       25.985**    36.462*     78.912***
##                      (12.003)    (9.867)     (10.589)    (19.878)    (28.080)
##
## MUJER:CASADO          -97.510***   -95.476***  -103.953*** -96.991***   -84.692*
##                      (18.984)    (19.394)    (22.300)    (31.892)    (44.438)
##
## MUJER:Numerohijos12    4.616        -0.700      -14.241     -19.080     -38.788
##                      (13.597)    (13.138)    (14.771)    (23.826)    (33.769)
##
## Constant              -32.997      -18.479       1.793       51.702     144.804***
##                      (24.446)    (24.599)    (25.920)    (40.321)    (53.623)
## -----
## Observations           10,469        10,469        10,469        10,469        10,469
## =====
## Note:                                     *p<0.1; **p<0.05; ***p<0.01

```

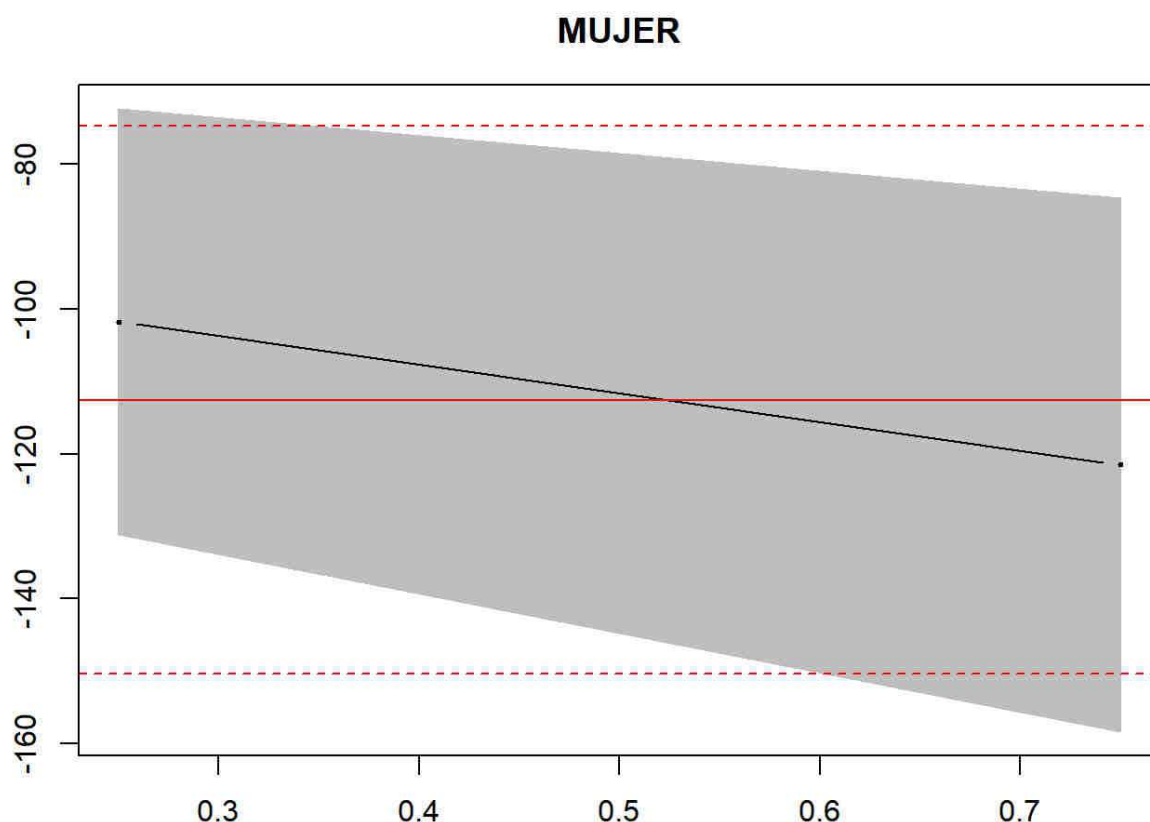
## 7 Diagn0sis sobre los efectos sobre los diferentes cuantiles

```
# ANOVA test for coefficient differences
anova(quantreg125,quantreg25,quantreg50, quantreg90)
```

```
## Warning in summary.rq(x, se = se, R = R, covariance = TRUE): 5 non-positive
## fis
```

```
## Quantile Regression Analysis of Deviance Table
##
## Model: SALARIO ~ MUJER + ESUPERIORES + ESECUNDARIOS + EXPERIENCIA + I(EXPERIENCIA^2) + ENFERMEDAD
+ HORASMENSUALES + SUPERVISOR + TEMPORAL + EXTRANJERO + CASADO + CASADO * MUJER + Numerohijos12 + Nu
merohijos12 * MUJER
## Joint Test of Equality of Slopes: tau in { 0.125 0.25 0.5 0.9 }
##
##   Df Resid Df F value    Pr(>F)
## 1 42    41834 44.886 < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
# Simultaneous quantile regression
quantreg2575 <- rq(modelo, data=datos, tau=c(0.25, 0.75))
sum_2575<-summary(quantreg2575, se="boot")
plot(sum_2575, parm=2,ols = TRUE,level = 0.95)
```



```
# Dibujamos los datos
quantreg.all <- rq(modelo, tau = seq(0.05, 0.95, by = 0.05), data=datos)
```

```
## Warning in rq.fit.br(x, y, tau = tau, ...): Solution may be nonunique
```

```
## Warning in rq.fit.br(x, y, tau = tau, ...): Solution may be nonunique
```

```
#quantreg.plot <- summary(quantreg.all, se="boot")  
quantreg.plot<-summary(rq(modelo,  
  tau = seq(0.05, 0.95, by = 0.05),  
  data=datos))
```

```
## Warning in rq.fit.br(x, y, tau = tau, ...): Solution may be nonunique
```

```
## Warning in rq.fit.br(x, y, tau = tau, ...): Solution may be nonunique
```

```
## Warning in summary.rq(xi, U = U, ...): 42 non-positive fis
```

```
## Warning in summary.rq(xi, U = U, ...): 2 non-positive fis
```

```
## Warning in summary.rq(xi, U = U, ...): 12 non-positive fis
```

```
## Warning in summary.rq(xi, U = U, ...): 2 non-positive fis
```

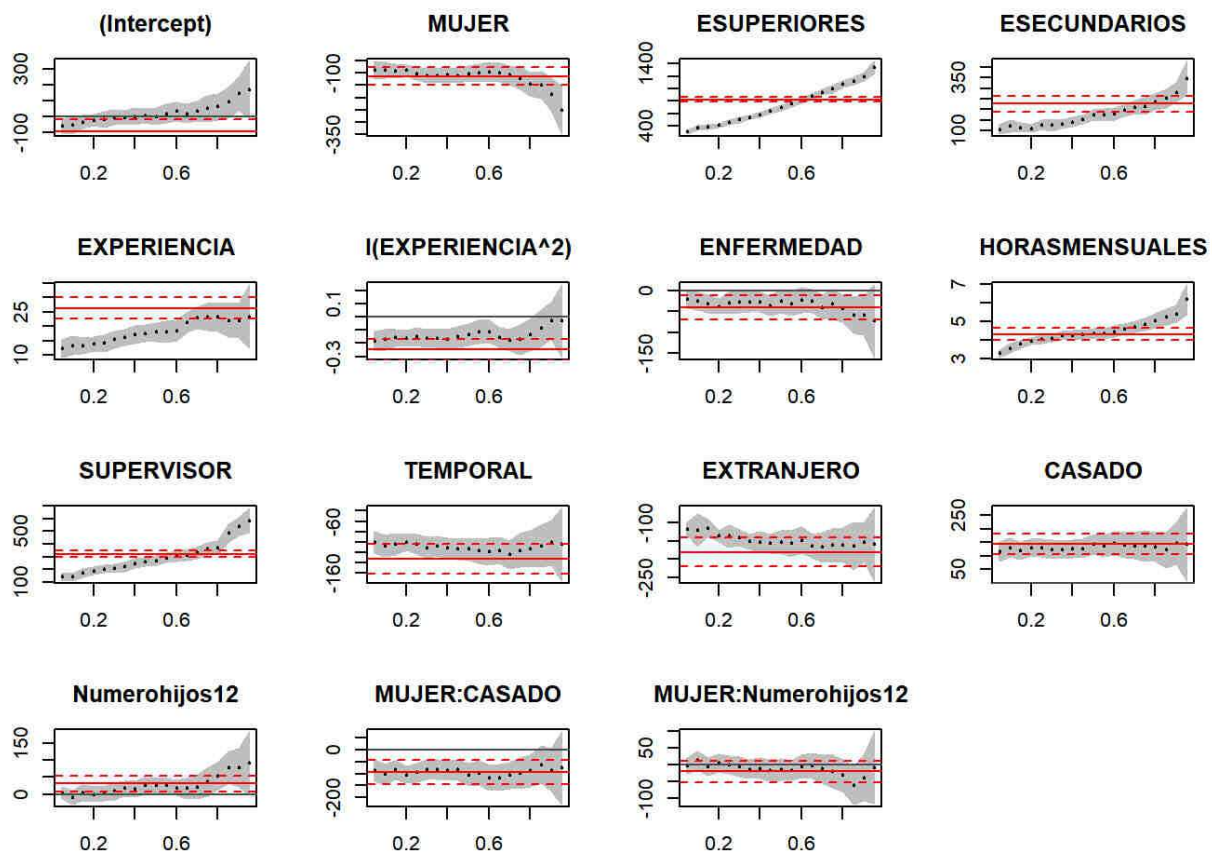
```
## Warning in summary.rq(xi, U = U, ...): 1 non-positive fis
```

```
## Warning in summary.rq(xi, U = U, ...): 1 non-positive fis
```

```
## Warning in summary.rq(xi, U = U, ...): 4 non-positive fis
```

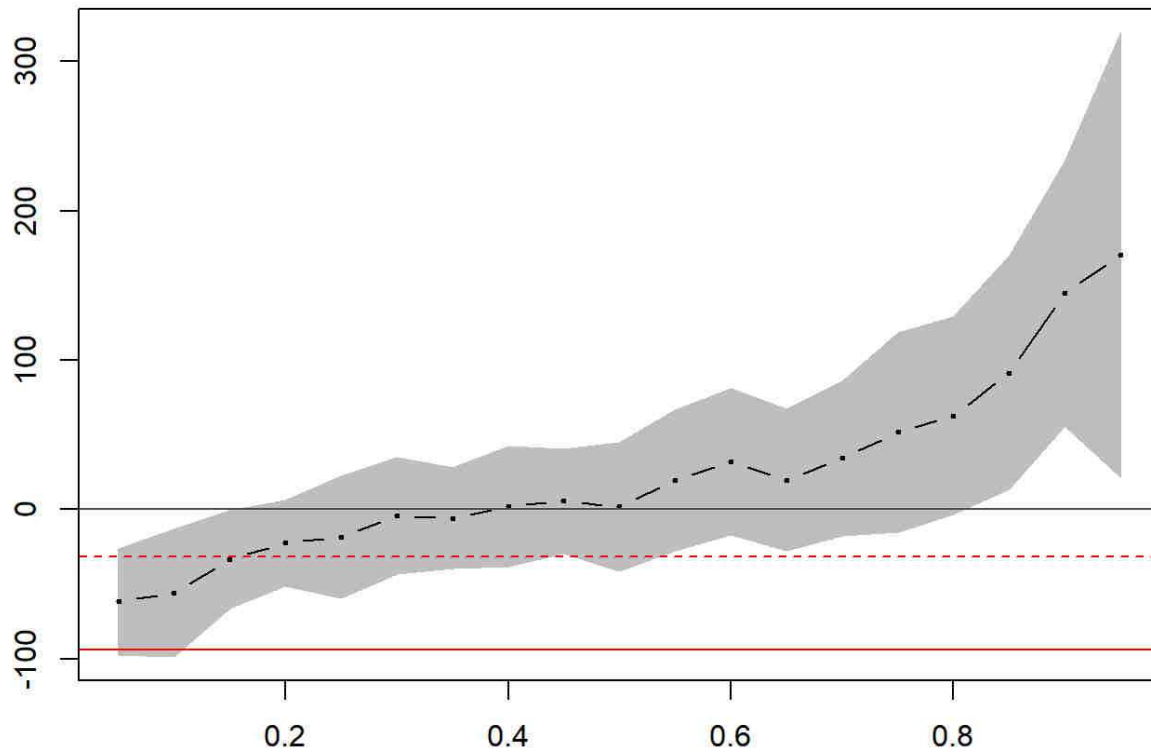
```
## Warning in summary.rq(xi, U = U, ...): 1 non-positive fis
```

```
plot(quantreg.plot,ols = TRUE, level=0.95)
```

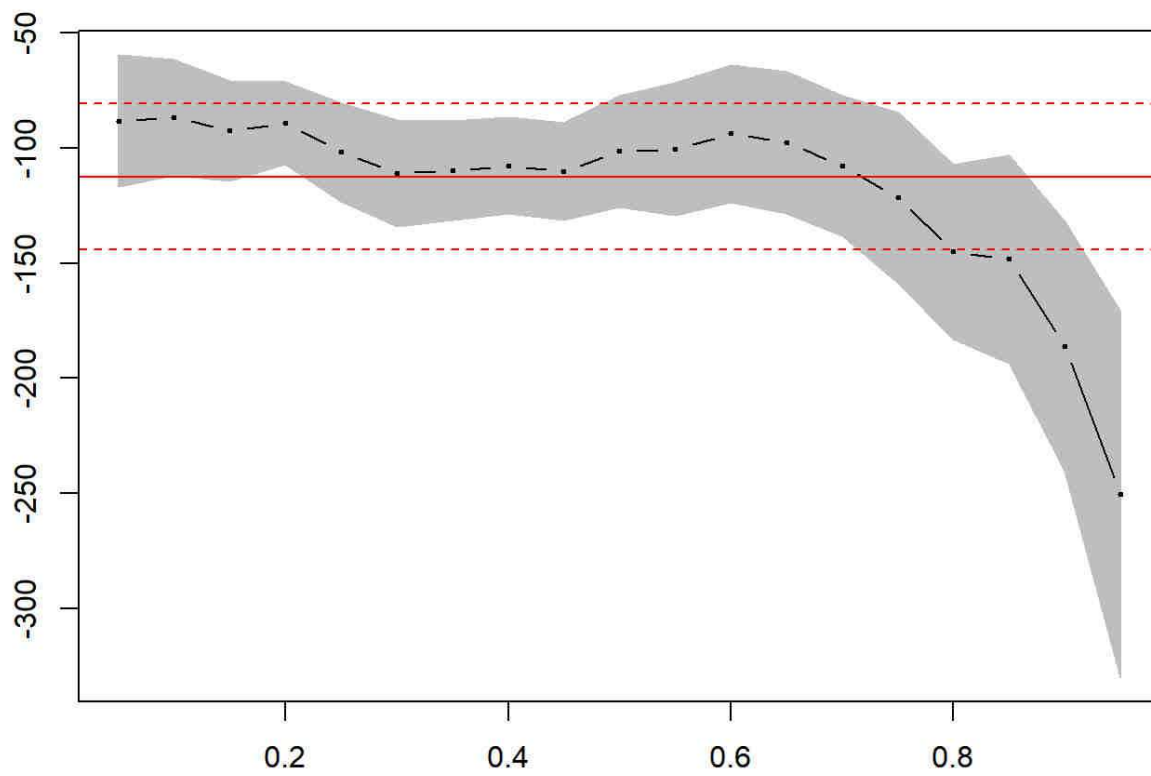


```
# puedo generar los gráficos para grabarlos en un pdf
# pdf("salidas_quantile_reg.pdf")
for( i in 1:15){
  plot(quantreg.plot, parm=i,ols=TRUE)
}
```

(Intercept)

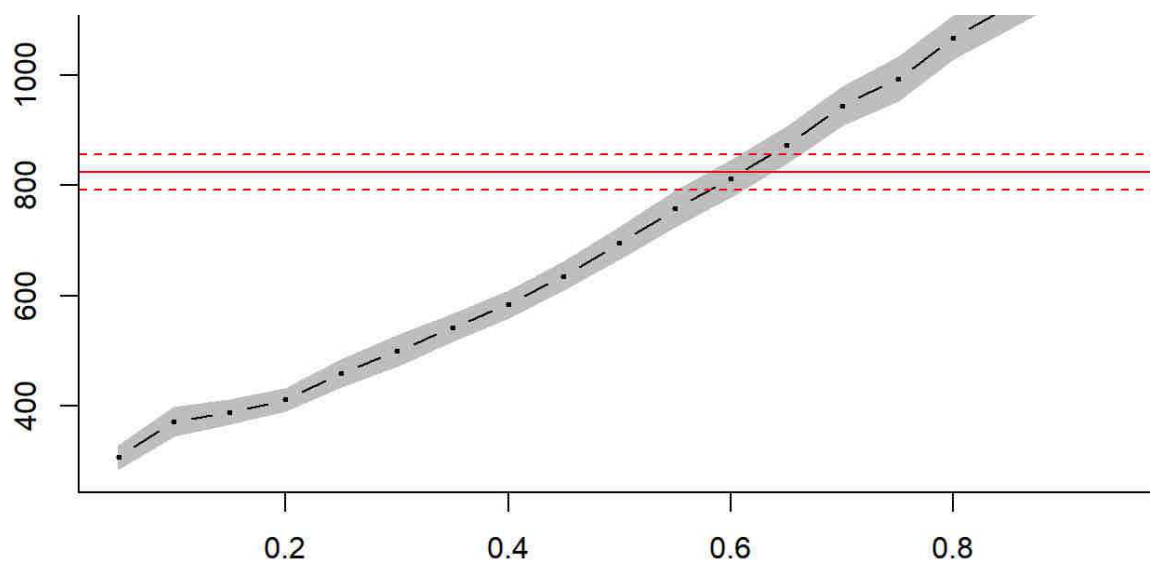


MUJER

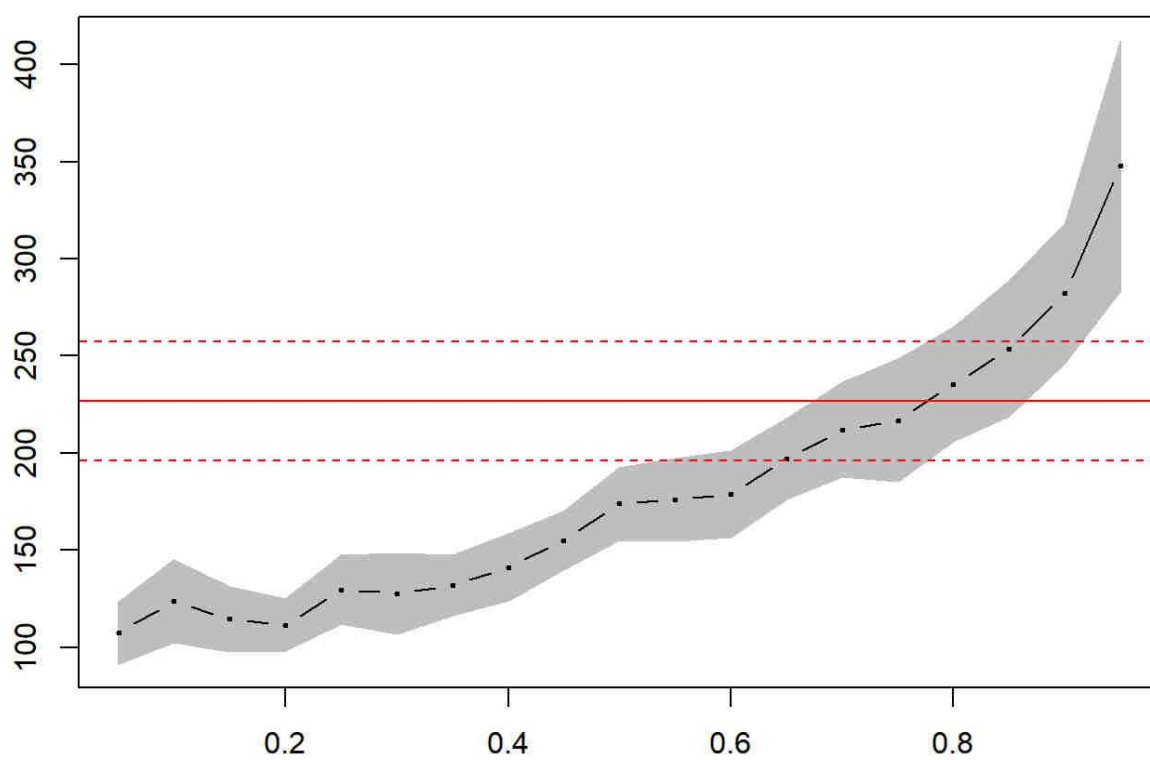


ESUPERIORES

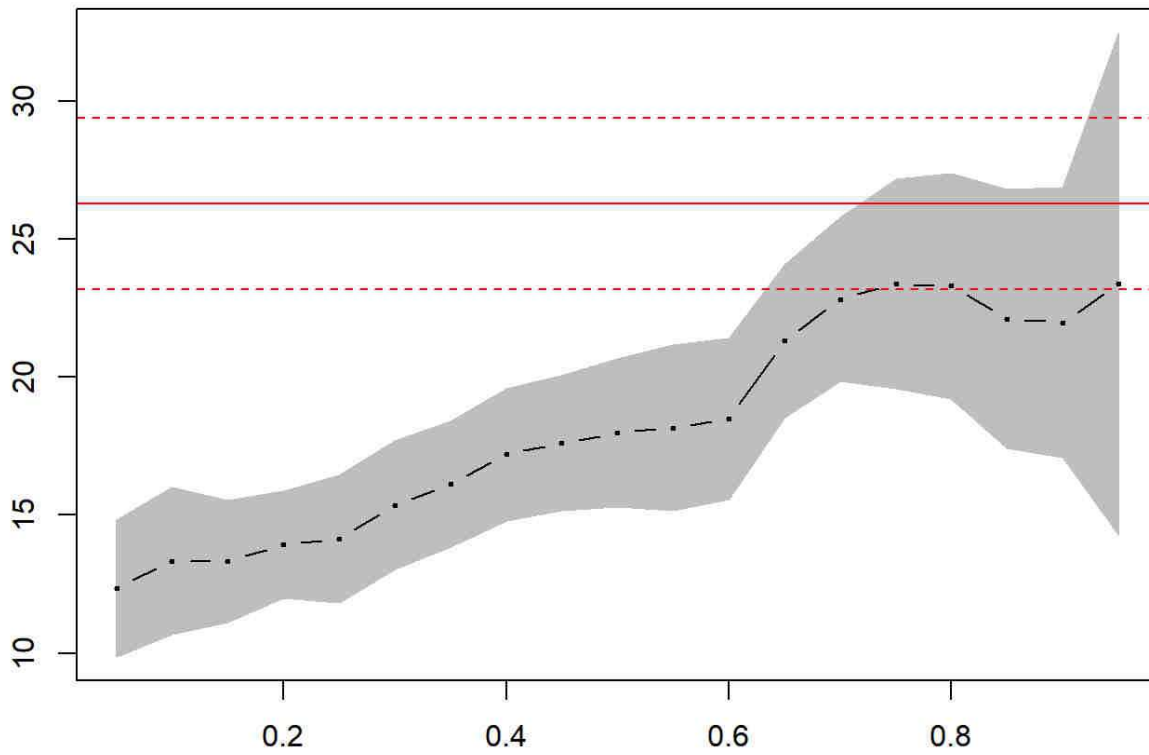




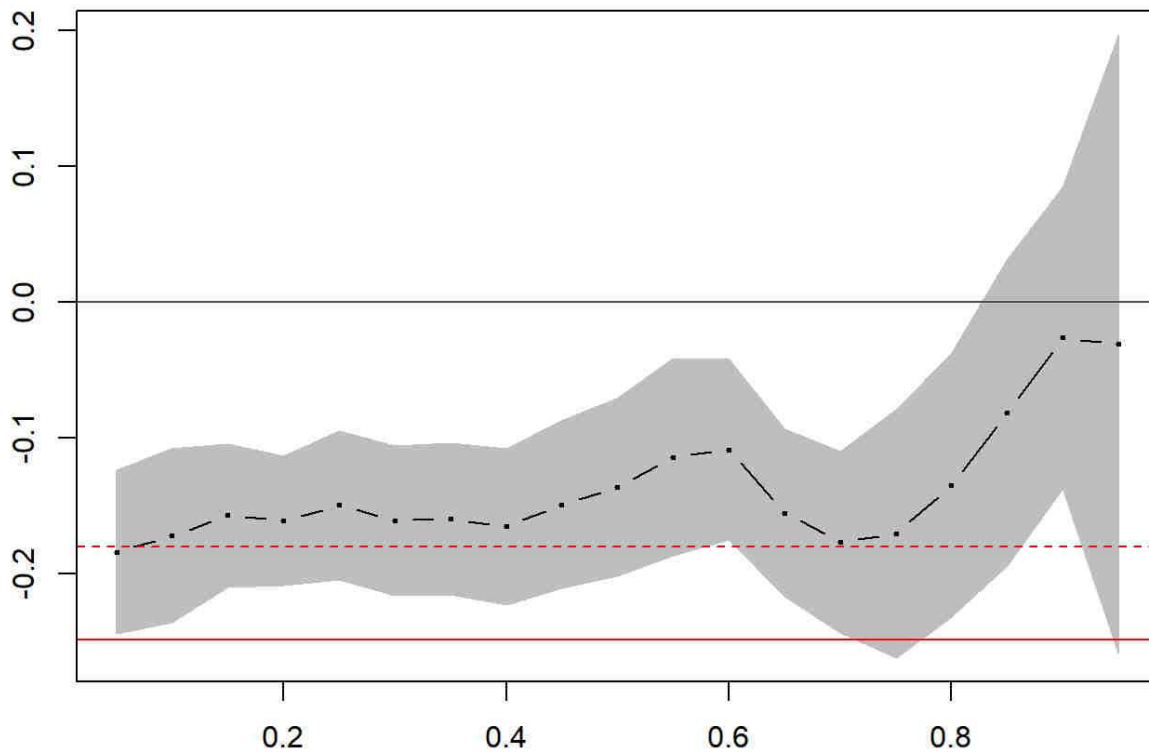
### ESECUNDARIOS



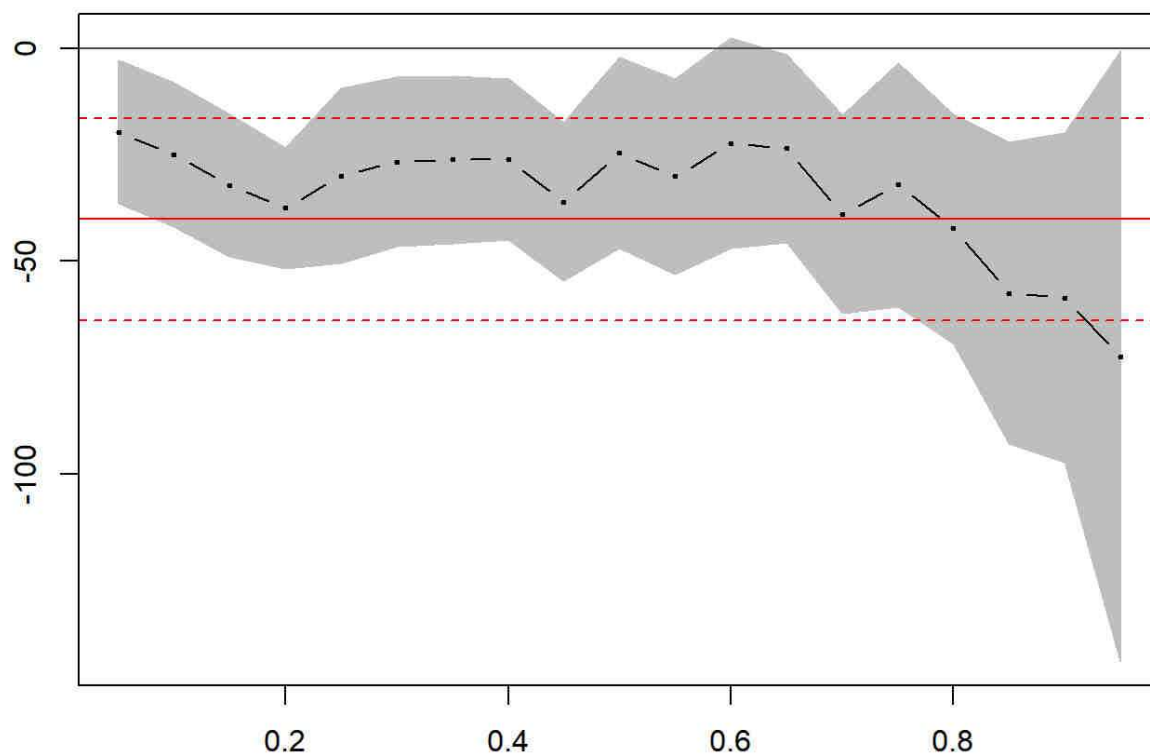
**EXPERIENCIA**



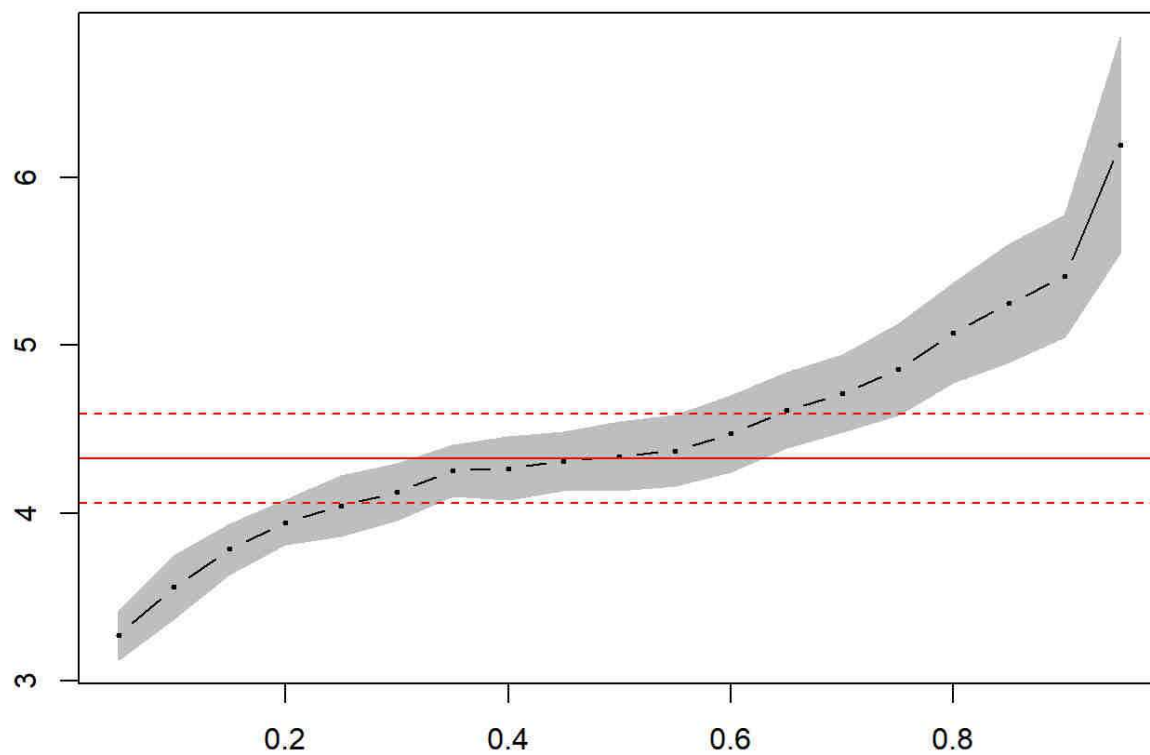
**I(EXPERIENCIA^2)**



## ENFERMEDAD

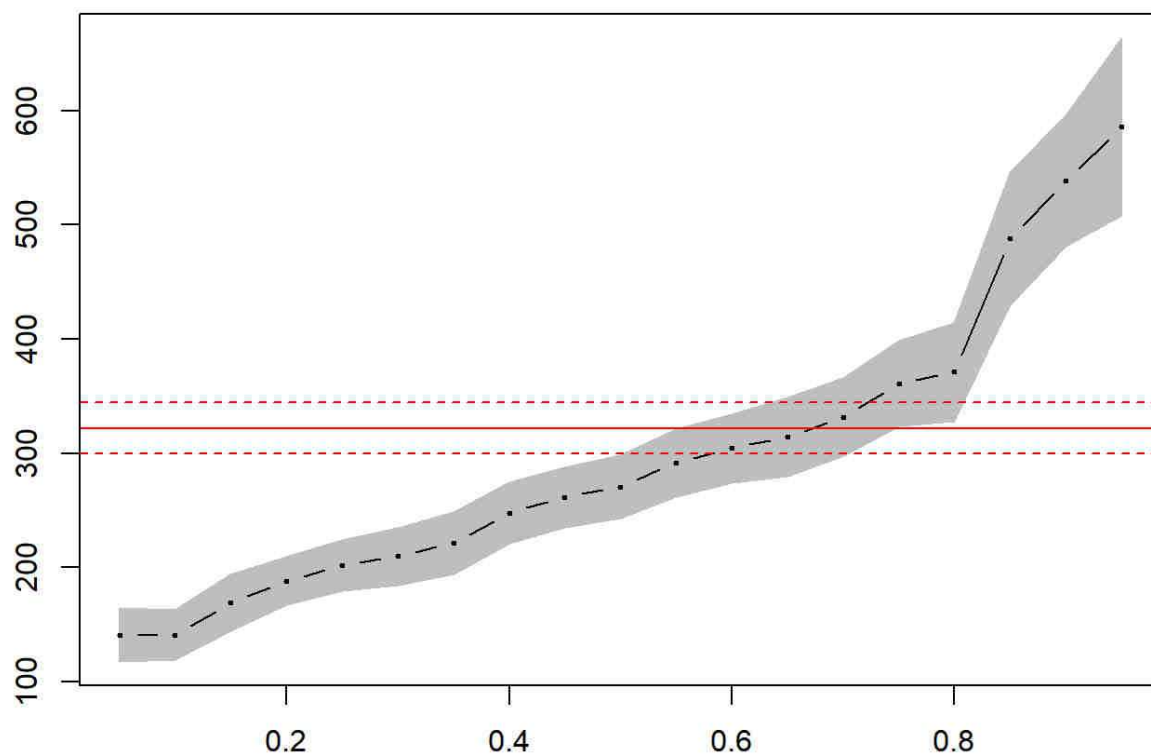


## HORASMENSUALES

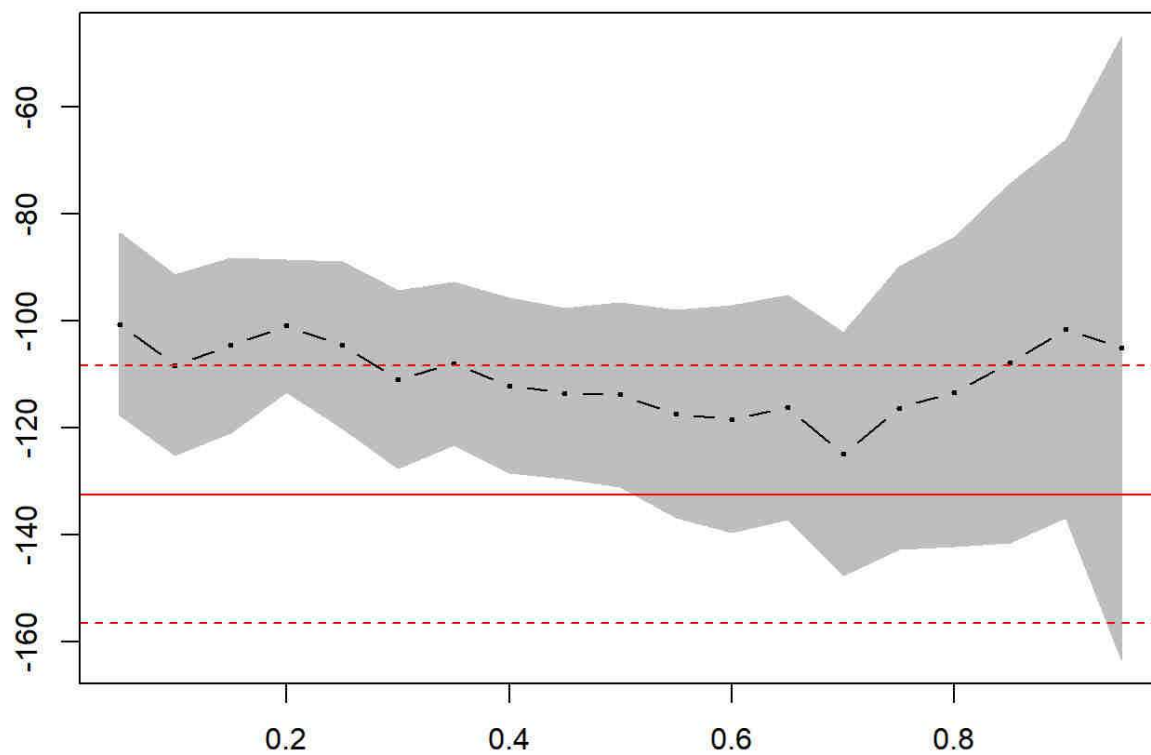




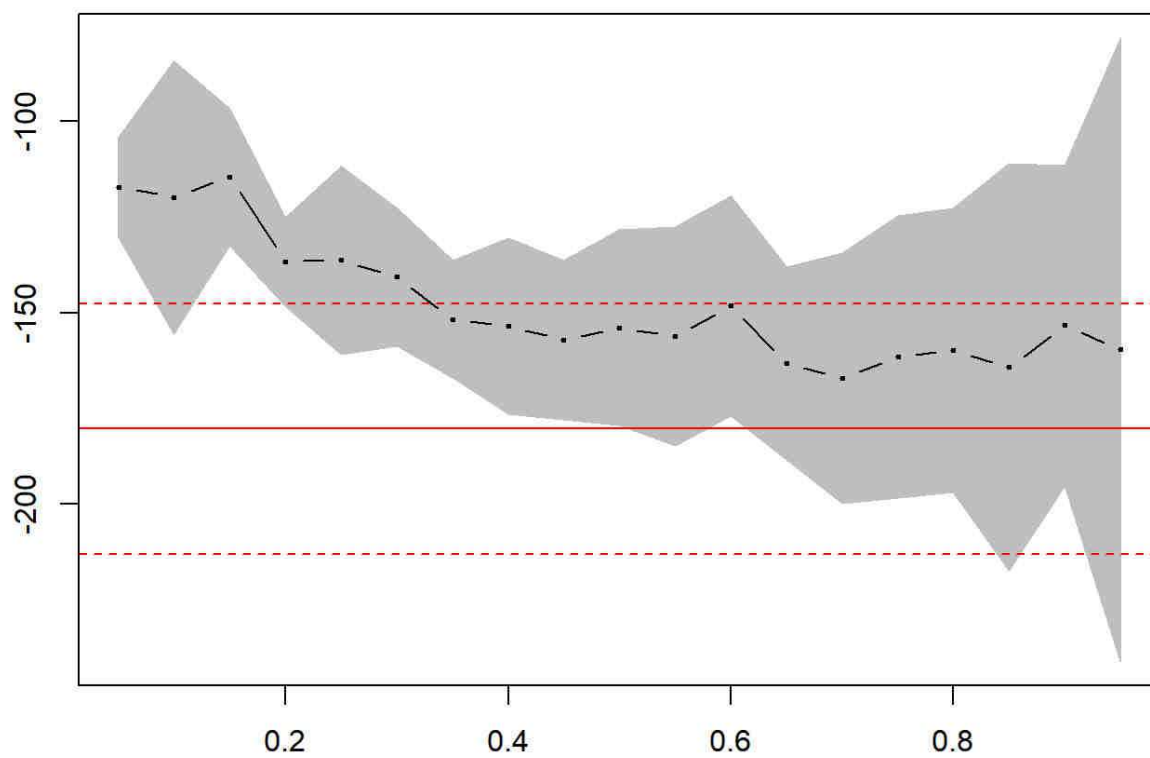
**SUPERVISOR**



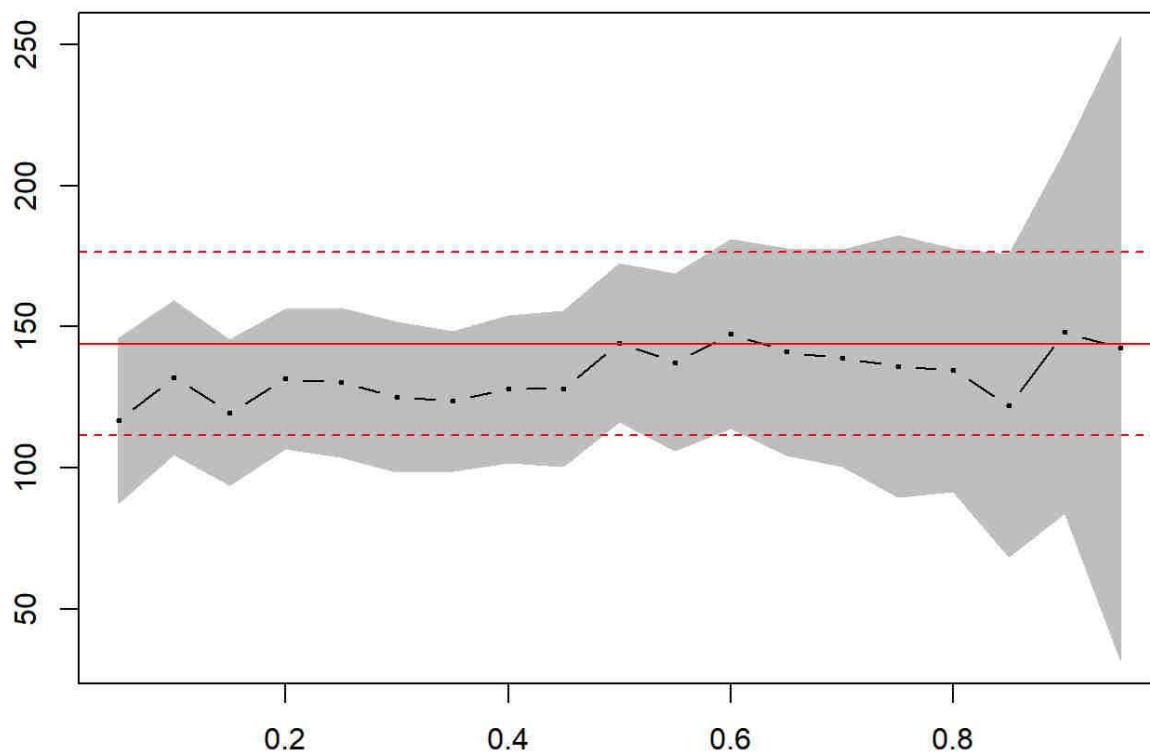
**TEMPORAL**



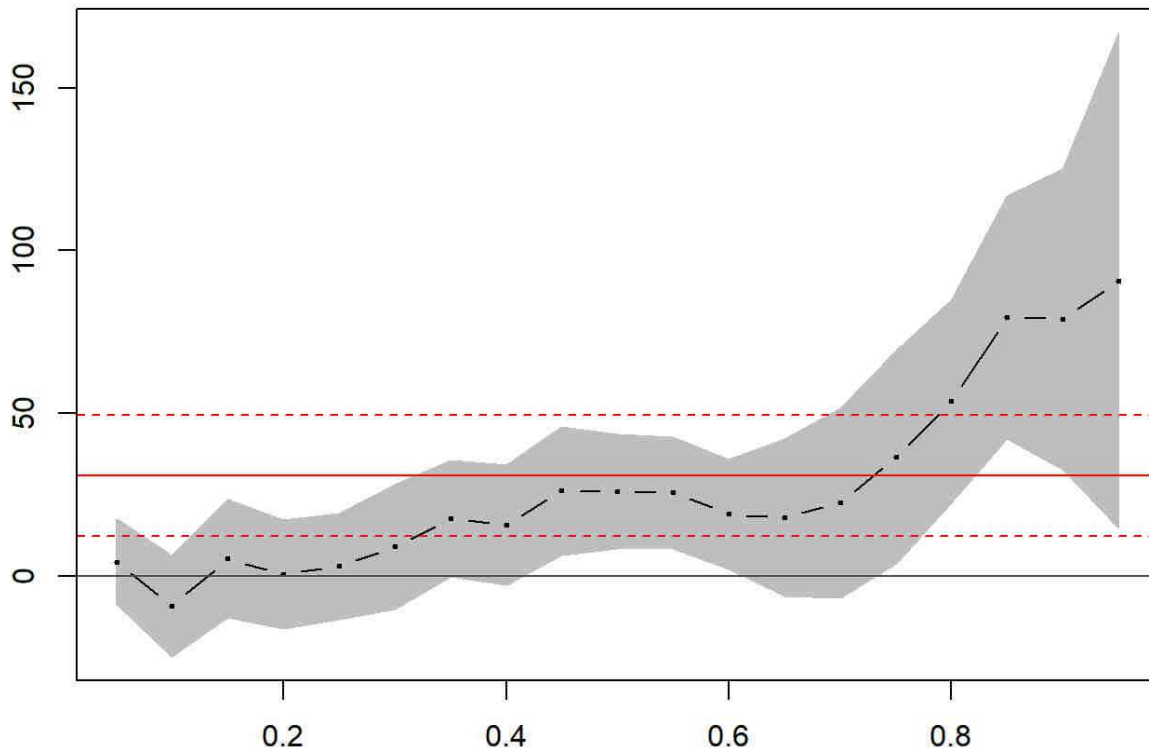
## EXTRANJERO



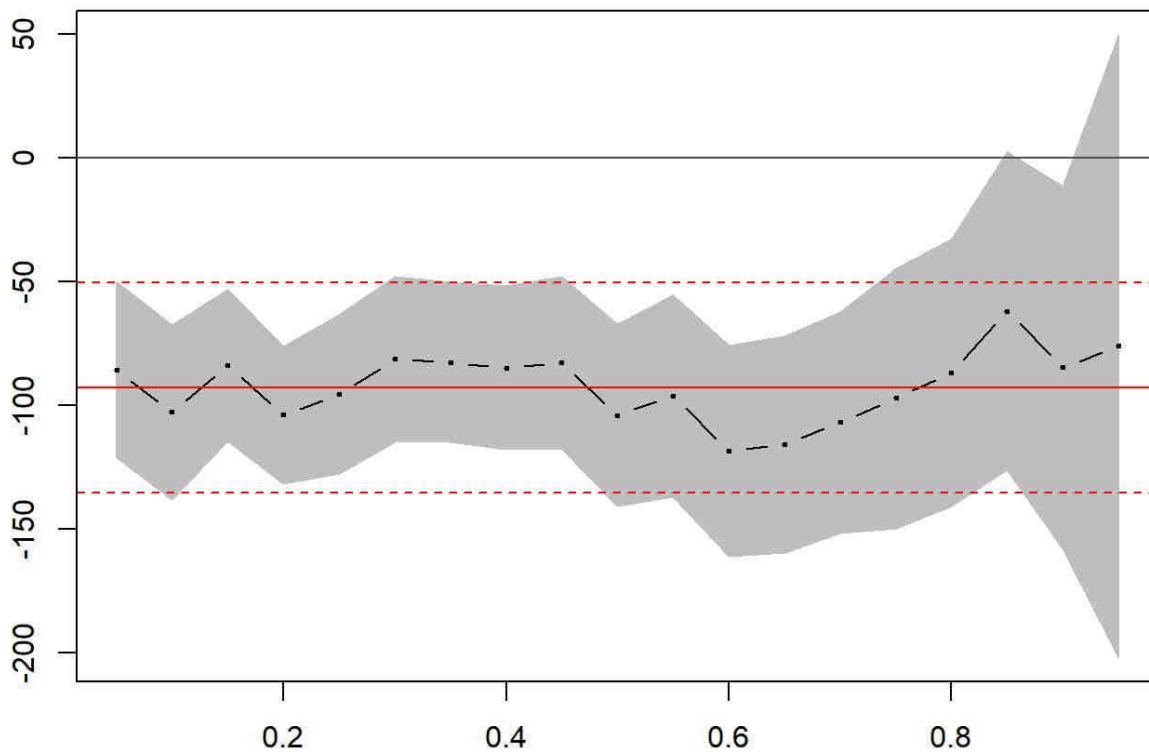
## CASADO



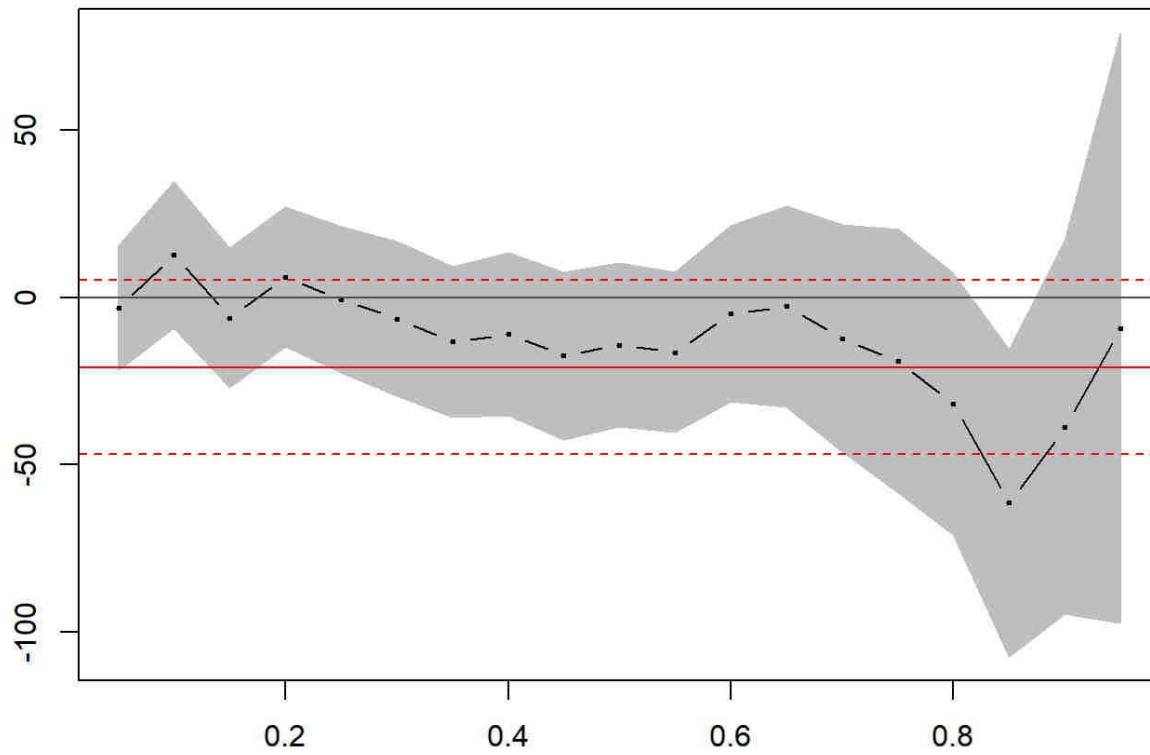
**Numerohijos12**



**MUJER:CASADO**



## MUJER: Numerohijos12



```
# si he guardado en un pdf tengo que cerrar  
# dev.off()
```



Proyecto de innovación  
“Innova Docencia” de la UCM  
Convocatoria 2018-2019  
Proyecto número 275

**R-adaptación de la asignatura de métodos econométricos en economía y finanzas del  
Grado de Estadística Aplicada**

Facultad de Estudios Estadísticos  
Departamento de Economía Aplicada, Pública y Política

**Título**

***Práctica 19 : Modelos de Econometría Espacial con R: La ley de Okun  
espacial***

**Correspondiente al**

**Capítulo 10 MODELOS DE ECONOMETRÍA ESPACIAL**

- Introducción a la estadística y a las medidas de asociación espacial
- La construcción de la matriz de distancias para la medición de dependencias espaciales
- Especificaciones del modelo con dependencia espacial: modelo de error con retardos espaciales y el modelo con errores espaciales autoregresivos
- Problemas de mínimos cuadrados ordinarios para la estimación de modelos con retardo espacial y el estimador de mínimos cuadrados en dos etapas

***Práctica 19 : Modelos de Econometría Espacial con R: La ley de Okun  
espacial***

# Modelos de Econometría Espacial

## Asignatura Métodos Econométricos en Economía y Finanzas

Profesores  
Loren Escot, Alicia Pérez-Alonso y Julio E. Sandubete

Dpto Economía Aplicada, Pública y Política  
Facultad Estudios Estadísticos - Universidad Complutense de Madrid

Curso académico 2018/2019

### Abstract

En esta práctica se presenta la estimación de modelos de regresión espacial con R. Se analizarán los datos de Desempleo y Renta Percápita a nivel de Municipios de España para analizar si en la estimación de la ley de Okun con datos georeferenciados existen efectos espaciales, y en tal caso, cómo estimarlos. Utilizaremos para ello cartografías, matrices de vecindad, test de correlación espacial, y la estimación de los modelos de error espacial y de retardo espacial.

- 1 Objetivo
- 2 Carga de Librerías necesaria
- 3 Carga de Cartografías de los Municipios Españoles
- 4 Mapas Descriptivos
  - 4.1 Por contiguidad de la Reina
  - 4.2 Por vecinos Cercanos
  - 4.3 Por distancias
- 5 Retardos Espaciales
- 6 Correlación entre una variable y su retardo espacial
  - 6.1 Coeficiente de Correlación espacial: La I de Moran **Global**
  - 6.2 Correlación Espacial **Local**
    - 6.2.1 Mapa de las correlaciones espaciales locales y significativas
- 7 Correlación entre Tasa de Paro y Renta Percápita
- 8 Modelo regresión Espacial
  - 8.1 Modelo sin efectos espaciales
  - 8.2 test sobre efectos espaciales en los residuos
  - 8.3 Test sobre los modelos espaciales
  - 8.4 Modelos de econometría Espacial
    - 8.4.1 SAR: Spatial lag Model
    - 8.4.2 SEM: Espatial Error Model
- 9 Otros Modelos

```
#=====
```

```
rm(list=ls()) # Elimina todos los objetos que tengamos en la memoria
```

## 1 Objetivo

Queremos estimar la ley de OKUN con datos Municipales, en este caso para explicar la Tasa de Paro Municipal en función de la renta percápita Municipal. Al tener datos georeferenciados nos preguntaremos si en la explicación de la tasa de paro municipal existen correlaciones espaciales con los municipios vecinos, y en tal caso presentaremos dos modelos para intentar estimar dichos efectos espaciales

## 2 Carga de Librerías necesaria

```
# suppressMessages(library(sp)) # para utilizar datos espaciales

suppressMessages(library(rgdal)) #para utilizar cartografías
suppressMessages(library(maptools)) # leer los shapes

suppressMessages(library(spdep)) # estadística espacial
suppressMessages(library(sphet)) # modelos de regresión espacial con heterocedasticidad
suppressMessages(library(RColorBrewer)) # paleta de colores
suppressMessages(library(classInt)) # para crear intervalos
suppressMessages(library(stargazer)) # tablas de resultados de modelos de regresión
suppressMessages(library(scales))
suppressMessages(library(dplyr)) # para manejar datos

suppressMessages(library(spatialreg)) # para otros métodos de regresión espacial
# suppressMessages(library(spse)) # para otros métodos de estimación
```

## 3 Carga de Cartografías de los Municipios Españoles

```
# Leo los Shapes con `maptools`
MUNIC_ESP<-maptools::readShapeSpatial("Munics04_GEO_ETRS89_RAUL.shp",IDvar = "cod_ine", proj4string=
  CRS("+init=epsg:4258"))
```

```
## Warning: readShapeSpatial is deprecated; use rgdal::readOGR or sf::st_read
```

```
## Warning: readShapePoly is deprecated; use rgdal::readOGR or sf::st_read
```

```
plot(MUNIC_ESP)
```



```
# Leo losshape con `rgdal`
#MUNIC_ESP<-readOGR( dsn= "." , layer="Munics04_GEO_ETRS89_RAUL")
#MUNIC_ESP<-readOGR( dsn= getwd() , layer="Munics04_GEO_ETRS89_RAUL")
#plot(MUNIC_ESP)

# El SpatialPoligonDataFrame que hemos cargado incorpora una base con datos Municipales
# Podemos consultar por ejemplo los municipios mñas ricos en la base de datos

#library(dplyr)
head(MUNIC_ESP@data%>%
      arrange(desc(RENTPCAP07))%>%
      select(MUN,RENTPCAP07), 10)
```

```
##              MUN RENTPCAP07
## 1           Madrid  40790.05
## 2          Barcelona  40178.55
## 3 Donostia/San Sebastián  39290.65
## 4           Zarautz  39242.60
## 5             Deyá  38578.88
## 6          Alcobendas  38525.45
## 7             Sitges  37308.56
## 8      Sant Just Desvern  35678.56
## 9             Getaria  34974.35
## 10  Sant Cugat del Vallès  34961.38
```

```
head(row.names(MUNIC_ESP@data), 10)
```

```
## [1] "01001" "01002" "01003" "01004" "01006" "01008" "01009" "01010"
## [9] "01011" "01013"
```



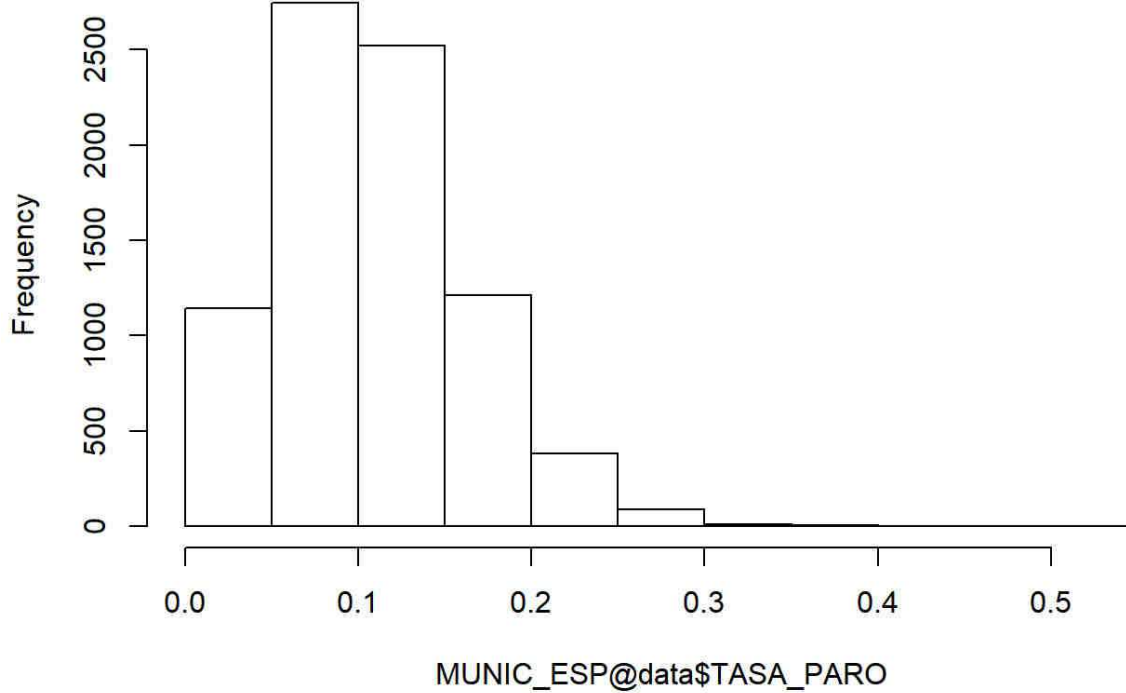
```
head(MUNIC_ESP@data,10)
```

##	SP_ID	id	cod_line	CODINE_MU	CCAA	COD_PROV	PROV
##	01001	01001	01001	01001	País Vasco	01	Araba/Álava
##	01002	01002	01002	01002	País Vasco	01	Araba/Álava
##	01003	01003	01003	01003	País Vasco	01	Araba/Álava
##	01004	01004	01004	01004	País Vasco	01	Araba/Álava
##	01006	01006	01006	01006	País Vasco	01	Araba/Álava
##	01008	01008	01008	01008	País Vasco	01	Araba/Álava
##	01009	01009	01009	01009	País Vasco	01	Araba/Álava
##	01010	01010	01010	01010	País Vasco	01	Araba/Álava
##	01011	01011	01011	01011	País Vasco	01	Araba/Álava
##	01013	01013	01013	01013	País Vasco	01	Araba/Álava
##		MUN	POB_2016	POB_HO	POB_MU	PARO_TO	PARO_CONS
##	01001	Alegría-Dulantzi	2856	1467	1389	195.83	17.08
##	01002	Amurrio	10260	5095	5165	856.16	89.08
##	01003	Aramaio	1502	792	710	28.58	0.41
##	01004	Artziniega	1840	911	929	114.58	5.58
##	01006	Armiñón	234	121	113	13.91	2.33
##	01008	Arratzua-Ubarrundia	979	531	448	39.33	0.91
##	01009	Asparrena	1630	821	809	106.33	6.83
##	01010	Ayala/Aiara	2946	1500	1446	140.00	10.00
##	01011	Baños de Ebro/Mañueta	308	168	140	7.75	0.00
##	01013	Barrundia	903	503	400	36.50	4.41
##	POB16_64	TASA_PARO	P_PAROCO	TASA_PA	RENTPCAP07	TA_PARO07	N_EDI
##	01001	1937	0.1010996	0.0872185	0.6782213	21456.95	3.41 667
##	01002	6836	0.1252428	0.1040460	0.6662768	22205.09	4.88 1316
##	01003	967	0.0295553	0.0143457	0.6438083	16977.71	1.70 423
##	01004	1225	0.0935347	0.0486996	0.6657609	19710.68	5.42 468
##	01006	167	0.0832934	0.1675054	0.7136752	15554.47	4.07 141
##	01008	697	0.0564275	0.0231376	0.7119510	24179.93	3.99 446
##	01009	1089	0.0976400	0.0642340	0.6680982	13268.75	3.17 532
##	01010	1896	0.0738397	0.0714286	0.6435845	19825.75	3.74 795
##	01011	194	0.0399485	0.0000000	0.6298701	9261.65	1.18 220
##	01013	600	0.0608333	0.1208219	0.6644518	15446.98	1.92 483
##	N_INM	N_IN_EDI	T_VIV_FA	T_VIV_PRI	P_VPRITOT	PrecioIn16	A_M_EDI
##	01001	1346	2.017991	1249	1082	0.8663029	1552.83 1993
##	01002	5154	3.916413	4435	3864	0.8712543	1665.67 1962
##	01003	706	1.669031	681	577	0.8473144	1252.81 NA
##	01004	1157	2.472222	1119	756	0.6757081	1476.80 NA
##	01006	153	1.085106	153	92	0.6009650	1165.40 NA
##	01008	484	1.085202	483	358	0.7412976	1757.80 NA
##	01009	1040	1.954887	981	677	0.6902528	979.12 NA
##	01010	1312	1.650314	1311	1071	0.8169588	1466.75 1945
##	01011	262	1.190909	240	119	0.4962558	683.43 NA
##	01013	512	1.060041	512	354	0.6914246	1139.86 NA
##	P_RUI	P_MALO	P_DEFI	P_BUE	FS_AGUA	FS_ELECT	FS_METEO
##	01001	0.0000000	0.0030075	0.0135338	0.9834586	9	2 0
##	01002	0.0030769	0.0184615	0.0730769	0.9053846	11	1 0
##	01003	NA	NA	NA	NA	2	0 0
##	01004	NA	NA	NA	NA	2	0 0
##	01006	NA	NA	NA	NA	0	0 0
##	01008	NA	NA	NA	NA	0	0 0
##	01009	NA	NA	NA	NA	2	0 0
##	01010	0.0000000	0.0302267	0.0806045	0.8891688	5	1 1
##	01011	NA	NA	NA	NA	0	0 1
##	01013	NA	NA	NA	NA	3	0 0
##	FS_INCEN	FS_OTROS	FS_CATAST	FS_ROBO	FS_LUNAS	FS_TOTAL	MD_AGUA
##	01001	0	11	0	0	0	22 29.44444
##	01002	3	2	0	2	2	39 99.72727
##	01003	0	0	0	0	0	2 22.50000
##	01004	0	7	0	0	0	9 0.50000
##	01006	0	1	0	0	0	1 0.00000

##	01008	0	2	0	1	2	5	0.00000
##	01009	0	1	0	0	1	4	7.00000
##	01010	0	12	0	0	2	21	3.00000
##	01011	0	1	0	0	0	2	0.00000
##	01013	0	1	0	0	0	4	14.00000
##		MD_ELECT	MD_METEO	MD_INCEND	MD_OTROS	MD_CATAST	MD_ROBO	MD_LUNAS
##	01001	33	0	0.000000	29.72727	0	0	0.0
##	01002	14	0	1.666667	56.30000	0	3	35.5
##	01003	0	0	0.000000	0.00000	0	0	0.0
##	01004	0	0	0.000000	3.20000	0	0	0.0
##	01006	0	0	0.000000	0.00000	0	0	0.0
##	01008	0	0	0.000000	1.00000	0	4	11.5
##	01009	0	0	0.000000	15.00000	0	0	57.0
##	01010	0	17	0.000000	26.75000	0	0	73.0
##	01011	0	18	0.000000	73.00000	0	0	0.0
##	01013	0	0	0.000000	6.00000	0	0	0.0
##		MD_TOTAL	PS_AGUA	PS_ELECT	PS_METEO	PS_INCEN	PS_OTROS	
##	01001	30.000000	0.0066865	0.0014859	0.0000000	0.0000000	0.0081724	
##	01002	59.461539	0.0021343	0.0001940	0.0000000	0.0005821	0.0003880	
##	01003	22.500000	0.0028329	0.0000000	0.0000000	0.0000000	0.0000000	
##	01004	2.428571	0.0017286	0.0000000	0.0000000	0.0000000	0.0060501	
##	01006	0.000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0065359	
##	01008	5.800000	0.0000000	0.0000000	0.0000000	0.0000000	0.0041322	
##	01009	44.000000	0.0019231	0.0000000	0.0000000	0.0000000	0.0009615	
##	01010	3.200000	0.0038110	0.0007622	0.0007622	0.0000000	0.0091463	
##	01011	45.500000	0.0000000	0.0000000	0.0038168	0.0000000	0.0038168	
##	01013	79.500000	0.0058594	0.0000000	0.0000000	0.0000000	0.0019531	
##		PS_CATAST	PS_ROBO	PS_LUNAS	PS_TOTAL			
##	01001	0	0.0000000	0.0000000	0.0163447			
##	01002	0	0.0003880	0.0003880	0.0075669			
##	01003	0	0.0000000	0.0000000	0.0028329			
##	01004	0	0.0000000	0.0000000	0.0077787			
##	01006	0	0.0000000	0.0000000	0.0065359			
##	01008	0	0.0020661	0.0041322	0.0103306			
##	01009	0	0.0000000	0.0009615	0.0038462			
##	01010	0	0.0000000	0.0015244	0.0160061			
##	01011	0	0.0000000	0.0000000	0.0076336			
##	01013	0	0.0000000	0.0000000	0.0078125			

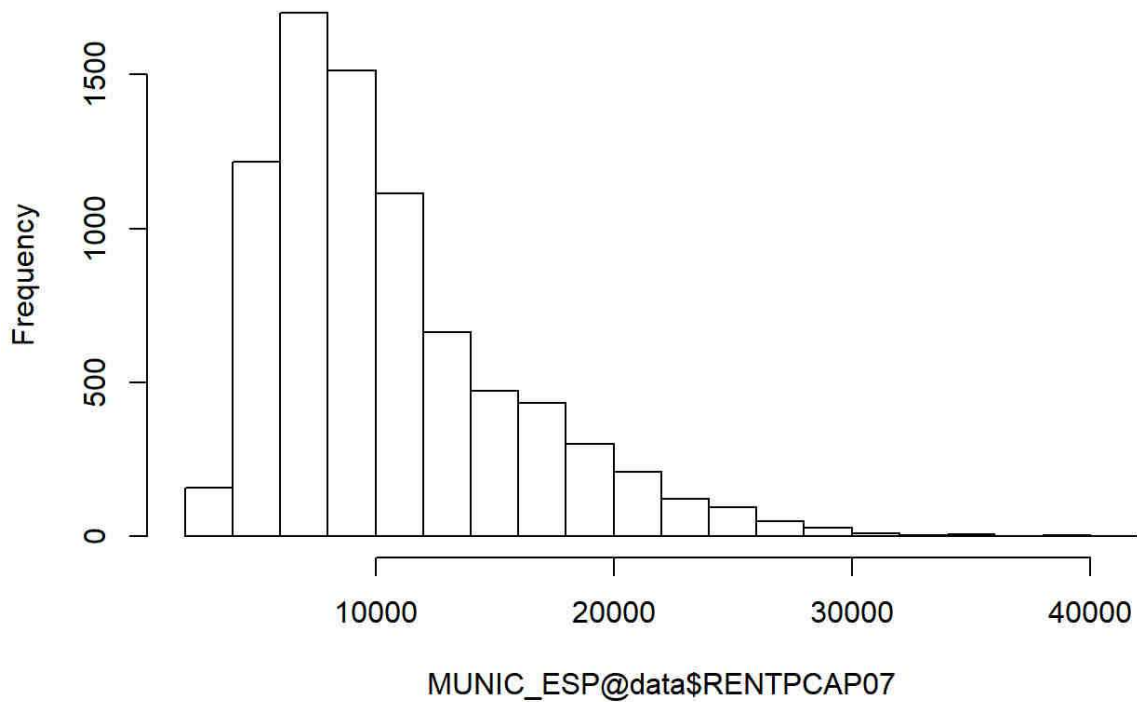
```
hist(MUNIC_ESP@data$TASA_PARO)
```

**Histogram of MUNIC\_ESP@data\$TASA\_PARO**



```
hist(MUNIC_ESP@data$RENTPCAP07)
```

**Histogram of MUNIC\_ESP@data\$RENTPCAP07**



```
# compruebo si hay datos nulos
```

```
table(is.na(MUNIC_ESP@data$RENTPCAP07))
```

```
##
## FALSE
## 8108
```

```
table(is.na(MUNIC_ESP@data$TASA_PARO))
```

```
##
## FALSE
## 8108
```

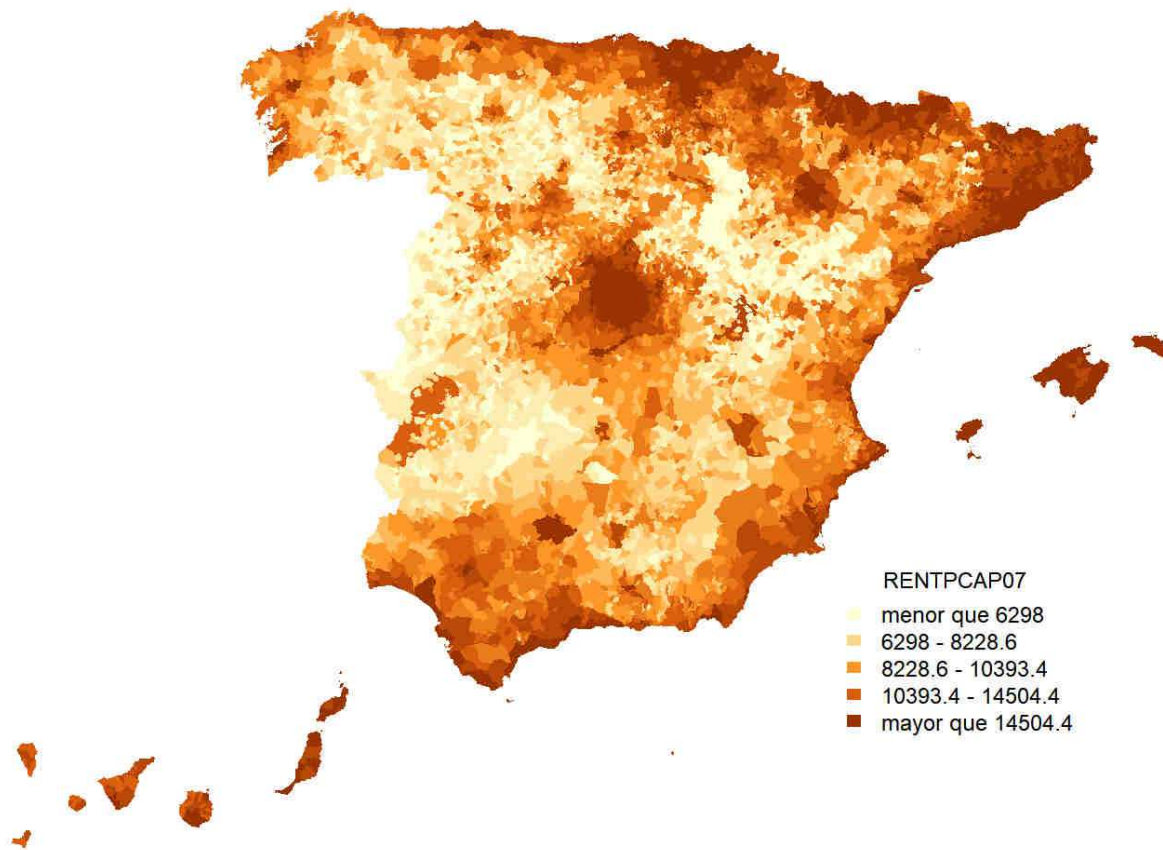
```
# si necesitase añadir algún campo adicional
# MUNIC_ESP@data<-dplyr::left_join(MUNIC_ESP@data,tablanewdatos, by=c("CODINE"="geo_code"))
# o tambien con sp::spCbind(MUNIC_ESP@data, tablanewdatos) los rownames deben ser los mismos
```

## 4 Mapas Descriptivos

```
# busco las coordenadas máximas y mínimas para poder representar los mapas
xmax<-max(coordinates(MUNIC_ESP)[,1])
xmin<-min(coordinates(MUNIC_ESP)[,1])

ymax<-max(coordinates(MUNIC_ESP)[,2])
ymin<-min(coordinates(MUNIC_ESP)[,2])

# Mapa de la renta per cápita por cuantiles
ColorBr="YlOrBr"
nombrevbles<- "RENTPCAP07"
par(mar=c(0,0,0,0))
breaks<-classIntervals(eval(parse(text=paste("MUNIC_ESP$",noquote(nombrevbles),sep=""))), 9, style="quantile")
color <- findColours(breaks,palette(brewer.pal(5, ColorBr)))
leg <- findColours(classIntervals(round(eval(parse(text=paste("MUNIC_ESP$",noquote(nombrevbles),sep=""))), digits=0), 5, style="quantile"),
                    palette(brewer.pal(5, ColorBr)),under="menor que", over="mayor que", between="-",
                    digits=0,cutlabels=FALSE)
#lo repito otra vez porque me falla a la primera No debería pero...
color <- findColours(breaks,palette(brewer.pal(5, ColorBr)))
leg <- findColours(classIntervals(round(eval(parse(text=paste("MUNIC_ESP$",noquote(nombrevbles),sep=""))), digits=0), 5, style="quantile"),
                    palette(brewer.pal(5, ColorBr)),under="menor que", over="mayor que", between="-",
                    digits=0,cutlabels=FALSE)
#dibujo el mapa
plot(MUNIC_ESP,col = color,lty=1, border=NA,lwd=0.25, add=F, axes=F, pch = 19,xlim=c(xmin,xmax), ylim=c(ymin,ymax))
legend(x = -0.70,y = 37.5,fill=attr(leg, "palette"),
       legend=names(attr(leg,"table")),
       title = nombrevbles, cex=0.75, box.lty=0, border = 0)
```

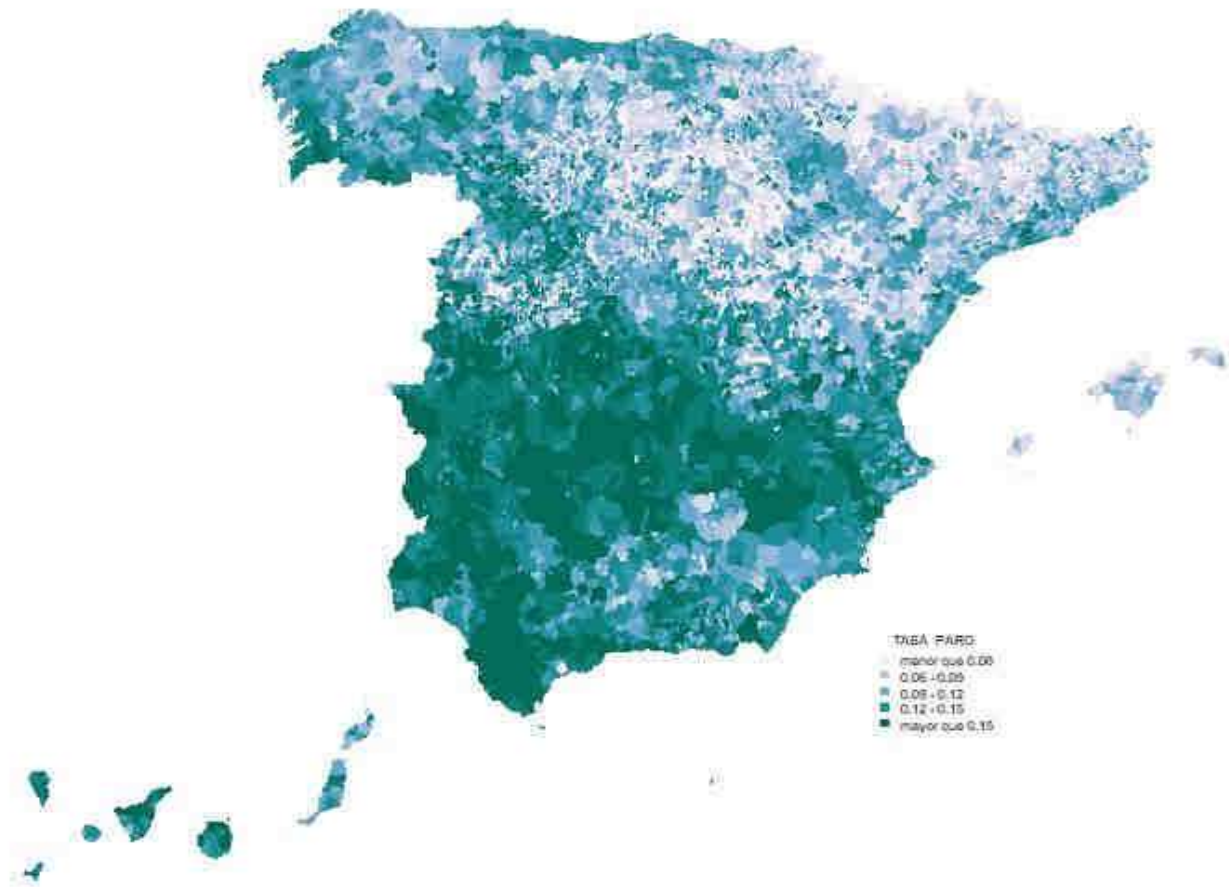


```
# Mapa de la Tasa de Paro
#Para grabarlo
jpeg(file=paste0("graficoTasa de Paro.jpeg"),height=500, width = 500*1.397,bg ="transparent", qualit
y = 100)
ColorBr="PuBuGn"
nombrevbles<-names(MUNIC_ESP@data)[15]
par(mar=c(0,0,0,0))
breaks<-classIntervals(eval(parse(text=paste("MUNIC_ESP$",noquote(nombrevbles),sep=""))), 9, style=
"quantile")
color <- findColours(breaks,palette(brewer.pal(5, ColorBr)))
leg <- findColours(classIntervals(round(eval(parse(text=paste("MUNIC_ESP$",noquote(nombrevbles),sep=
""))), digits=2), 5, style="quantile"),
palette(brewer.pal(5, ColorBr)),under="menor que", over="mayor que", between="-",
digits=0,cutlabels=FALSE)
#lo repito para que coja bien los colores
color <- findColours(breaks,palette(brewer.pal(5, ColorBr)))
leg <- findColours(classIntervals(round(eval(parse(text=paste("MUNIC_ESP$",noquote(nombrevbles),sep=
""))), digits=2), 5, style="quantile"),
palette(brewer.pal(5, ColorBr)),under="menor que", over="mayor que", between="-",
digits=0,cutlabels=FALSE)
#mapa
plot(MUNIC_ESP,col = color,lty=1, border=NA,lwd=0.25, add=F, axes=F, pch = 19,xlim=c(xmin,xmax), yl
im=c(ymin,ymax))
legend(x = -0.75,y = 37,fill=attr(leg, "palette"),
legend=names(attr(leg,"table")),
title = nombrevbles, cex=0.6, box.lty=0, border = 0)

#Si lo he grabado
dev.off()
```

```
## png
## 2
```

```
par(mar=c(5.1, 4.1, 4.1, 2.1))
```



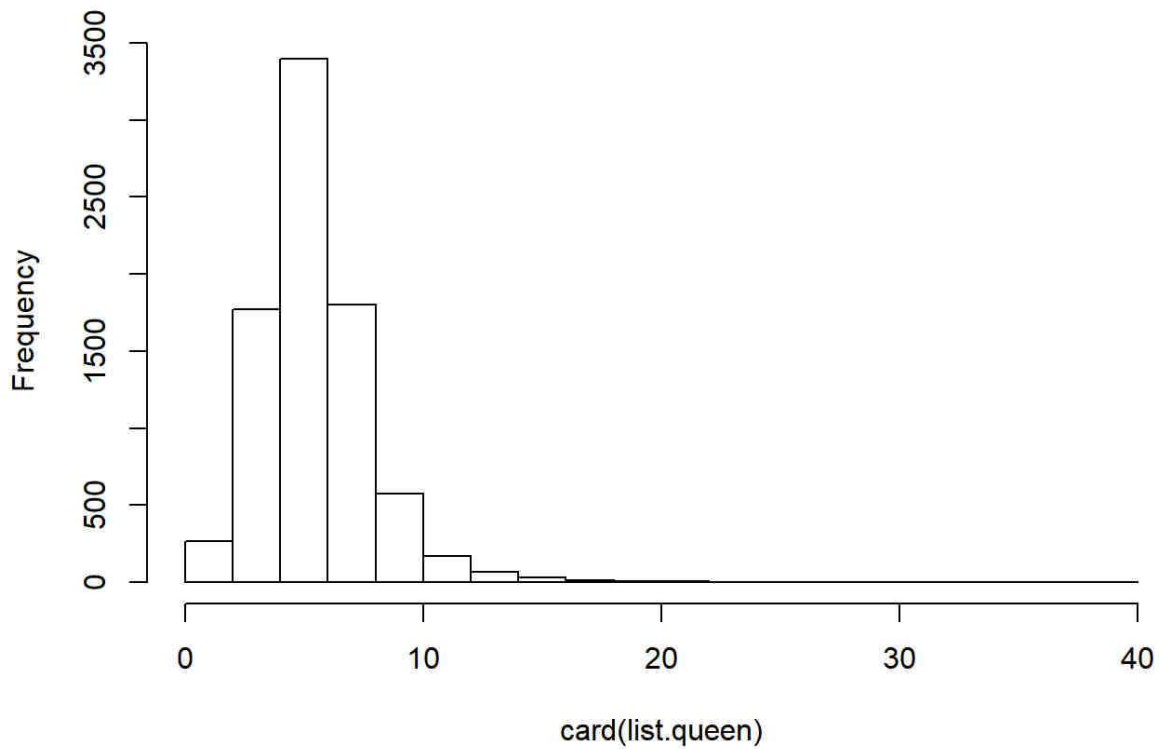
Observando esos dos gráficos, ¿podría decirse que los municipios con mayor nivel de Renta son los que tienen menores tasas de paro? # La matriz de pesos espaciales

```
#####  
# Creo la matriz de pesos
```

## 4.1 Por contigüidad de la Reina

```
#### Por contigüidad de la Reina  
list.queen<-poly2nb(MUNIC_ESP, queen=TRUE)  
W<-nb2listw(list.queen, style="W", zero.policy=TRUE)  
  
# histograma del número de vecinos  
hist(card(list.queen))
```

**Histogram of card(list.queen)**

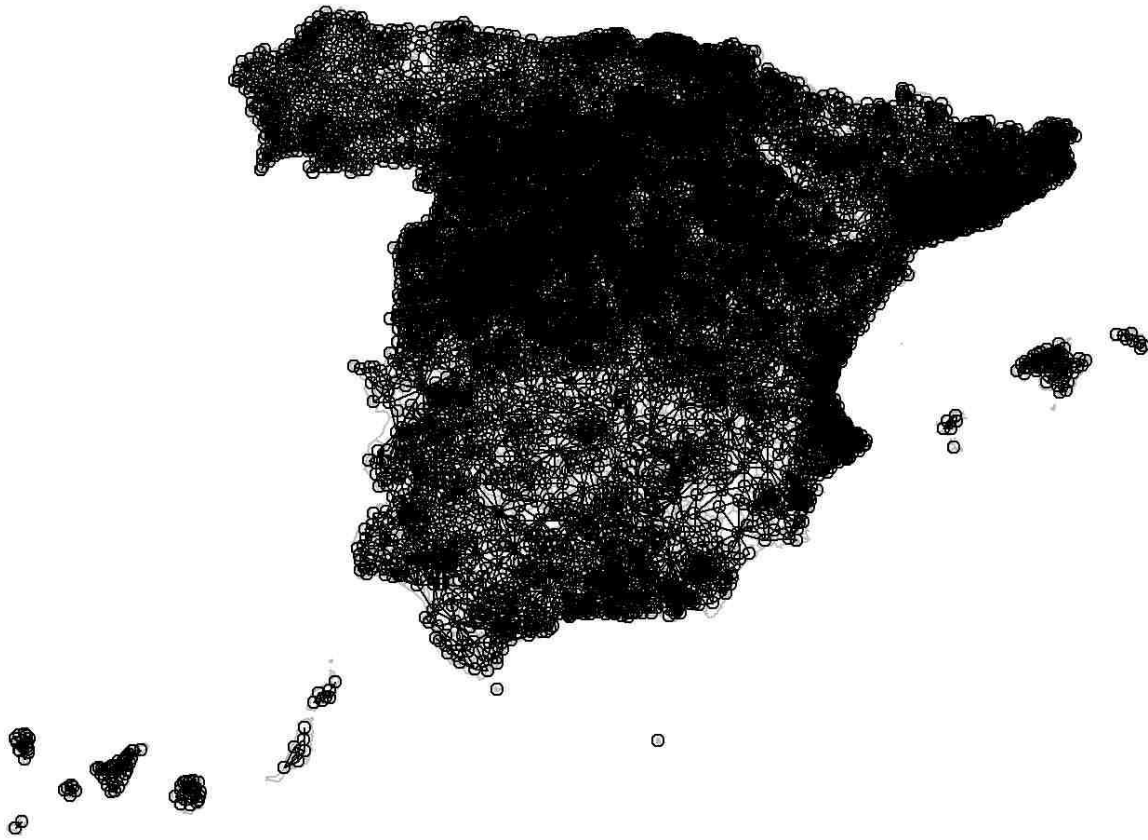


```
summary(card(list.queen))
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.000   4.000   6.000   5.964   7.000  39.000
```

```
# mapa de contigüidades
par(mar=c(0,0,0,0))
plot(MUNIC_ESP, border="grey")
plot(W,coordinates(MUNIC_ESP),add=TRUE)
```



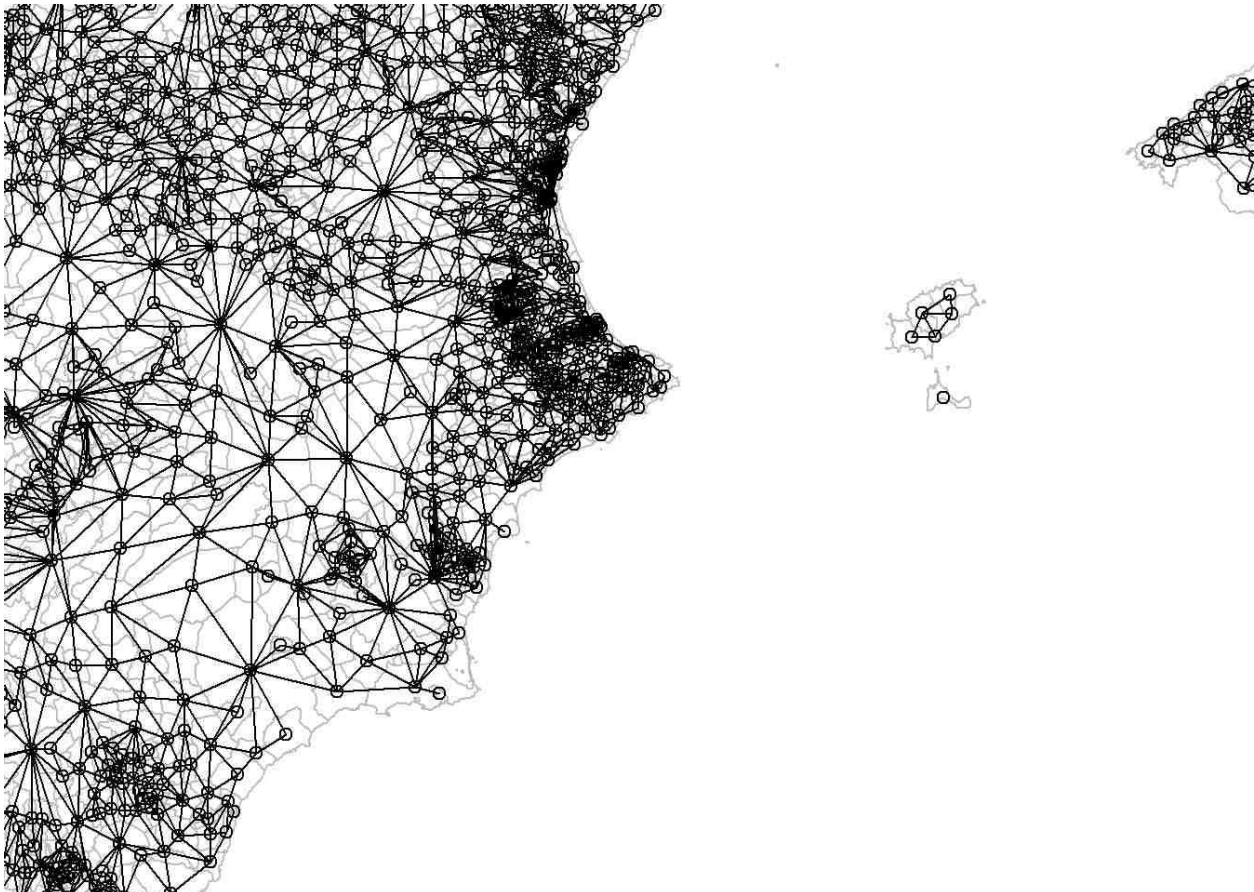


```
par(mar=c(5.1, 4.1, 4.1, 2.1))
```

```
par(mar=c(0,0,0,0))
```

```
plot(MUNIC_ESP, border="grey",xlim=c(-0.5,0.5), ylim=c(37,40))
```

```
plot(W,coordinates(MUNIC_ESP), add=TRUE)
```



```
par(mar=c(5.1, 4.1, 4.1, 2.1))
```

## 4.2 Por vecinos Cercanos

```
##### PARA ESTIMAR LA MATRIZ DE PESOS POR VECINOS CERCANOS
#coords <- coordinates(MUNIC_ESP)
#col.knn <- knearneigh(coords, k=4)

#par(mar=c(0,0,0,0))
#plot(MUNIC_ESP, border="grey", xlim=c(-0.5,0.5), ylim=c(37,40))
#plot(knn2nb(col.knn), coords, add=TRUE)
#title(main="K nearest neighbours, k = 4")
#par(mar=c(5.1, 4.1, 4.1, 2.1))

#Knn4<-knn2nb(col.knn)
#dists <- nbdists(Knn4, coords)
#W <- nb2listw(Knn4, glist=dists, style="W",zero.policy=TRUE)

## histograma de vecinos próximos
#hist(card(Knn4))
#summary(card(Knn4))

#par(mar=c(0,0,0,0))
#plot(MUNIC_ESP, border="grey",xlim=c(-0.5,0.5), ylim=c(37,40))
#plot(W,coordinates(MUNIC_ESP), add=TRUE)
#par(mar=c(5.1, 4.1, 4.1, 2.1))
```

## 4.3 Por distancias

```
##### PARA LA ESTIMACION DE LA MATRIZ DE PESOS POR DISTANCIA
#coords<-coordinates(MUNIC_ESP)

## Distancia entre 0 y la distancia máxima del más cercano (Asegura que todos tienen al menos uno)
#all.linked <- max(unlist(nbdists(knn2nb(knearneigh(coords)), coords)))
#W_dist<-dnearneigh(coords,0,all.linked,longlat = FALSE)

## Distancia entre 0 y 15 Kms
#W_dist<-dnearneigh(coords,0,15,longlat = TRUE) # distancia en Km

## histograma del número de vecinos
#hist(card(W_dist))

#par(mar=c(0,0,0,0))
#plot(MUNIC_ESP, border="grey",xlim=c(-0.5,0.5), ylim=c(37,40))
#plot(W_dist,coordinates(MUNIC_ESP), add=TRUE)
#par(mar=c(5.1, 4.1, 4.1, 2.1))

#dists <- nbdists(W_dist, coords)
#W <- nb2listw(W_dist, glist=dists, style="W",zero.policy=TRUE)

#par(mar=c(0,0,0,0))
#plot(MUNIC_ESP, border="grey",xlim=c(-0.5,0.5), ylim=c(37,40))
#plot(W,coordinates(MUNIC_ESP), add=TRUE)
#par(mar=c(5.1, 4.1, 4.1, 2.1))

#####
#####
# Una vez que he estimado W (lista) de la forma que nos parezca más adecuada
# Puedo convertirla en matriz para calcular la matriz de pesos y
# y los retardos espaciales

# Para este ejemplo me quedo con la matriz de vecidandes tipo REINA

# y ahora obtengo WM en forma de matriz para calcular retardos

WM<-listw2mat(W)
dim(WM)
```

```
## [1] 8108 8108
```

## 5 Retardos Espaciales

```
# Creo el retardo espacial de la tasa de paro
MUNIC_ESP@data$TASA_PARO<-(WM%*%MUNIC_ESP@data$TASA_PARO)[,1]

summary(MUNIC_ESP@data$TASA_PARO)
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.00000 0.06912 0.10230 0.10787 0.14115 0.52929
```

```
summary(MUNIC_ESP@data$TASA_PAROW)
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.00000 0.07600 0.10270 0.10910 0.13610 0.26950
```

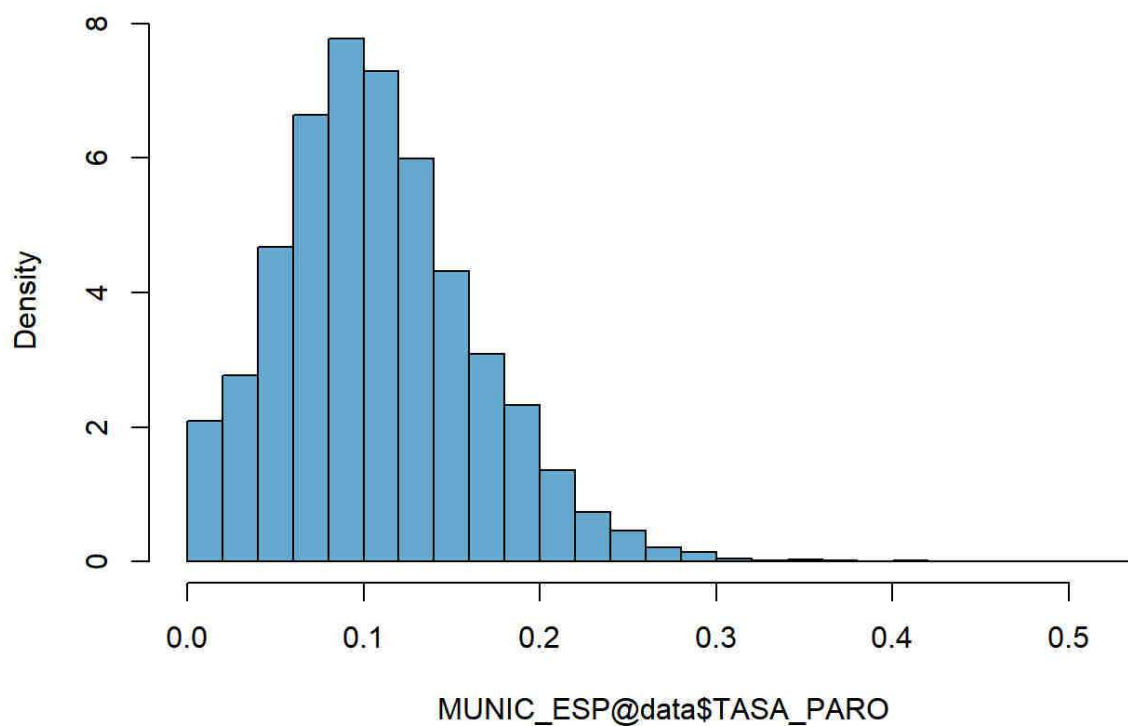
```
table(is.na(MUNIC_ESP@data$TASA_PAROW))
```

```
##
## FALSE
## 8108
```

```
par(mar=c(5.1, 4.1, 4.1, 2.1))
```

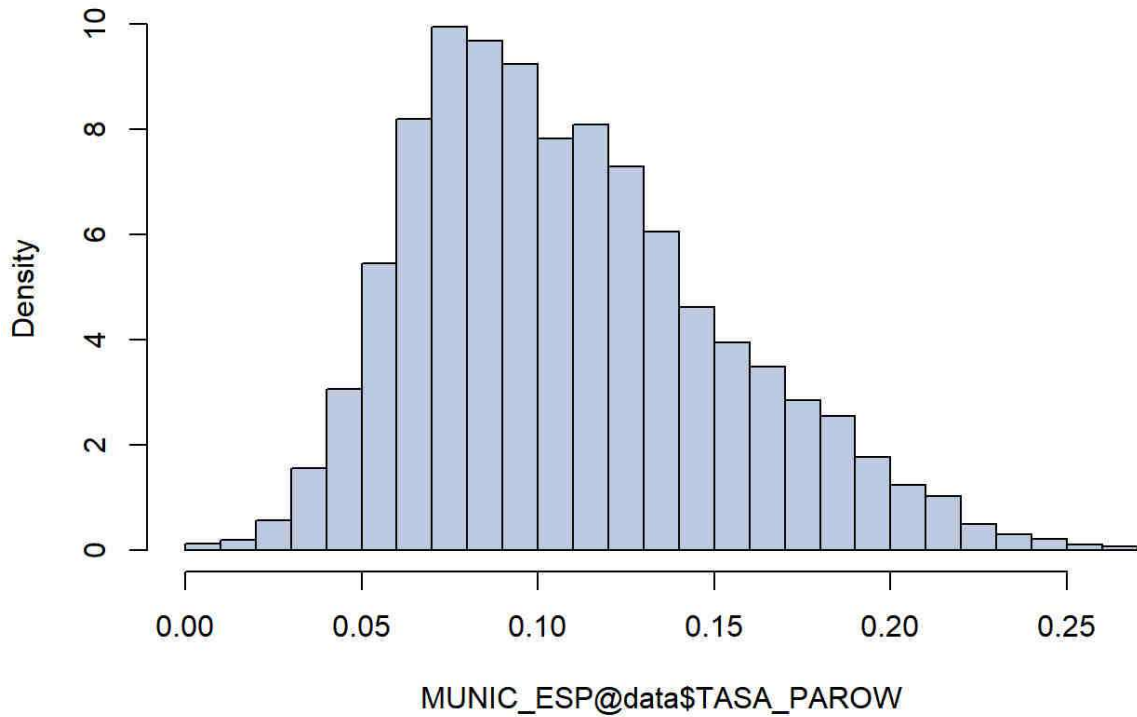
```
hist(MUNIC_ESP@data$TASA_PARO, nclass = 20, col=3, freq = FALSE)
```

**Histogram of MUNIC\_ESP@data\$TASA\_PARO**

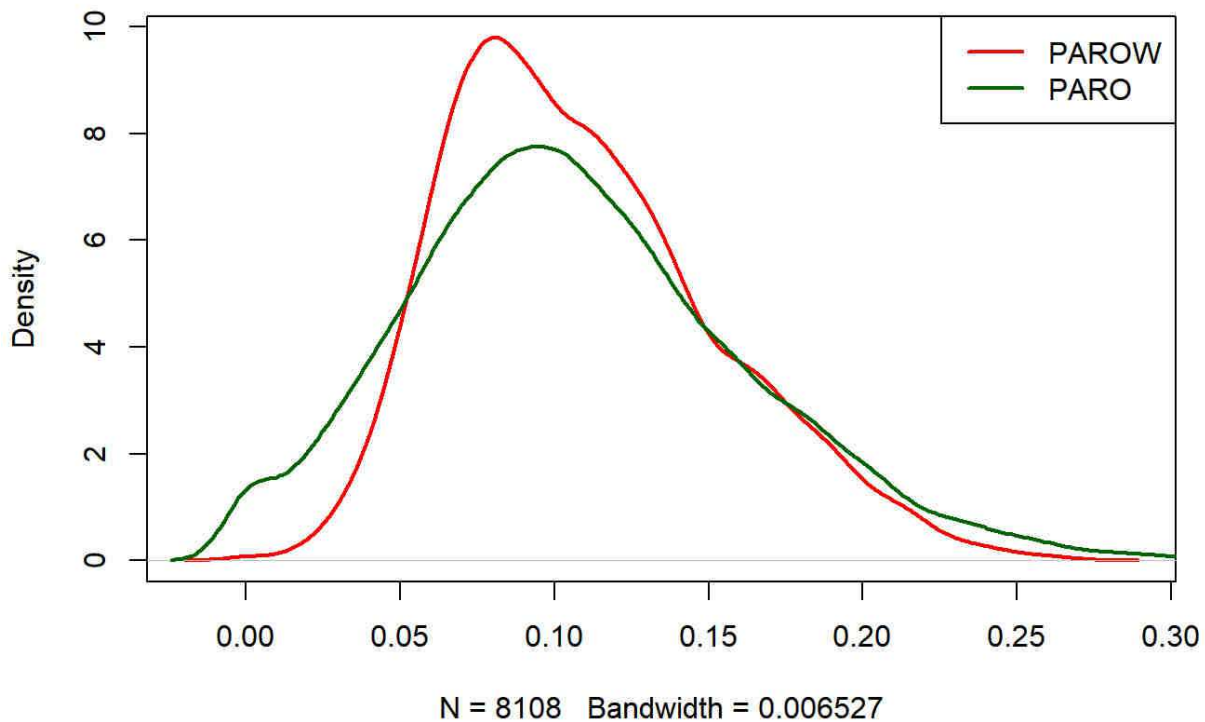


```
hist(MUNIC_ESP@data$TASA_PAROW, nclass=20, col=2, freq=FALSE)
```

**Histogram of MUNIC\_ESP@data\$TASA\_PAROW**



```
plot(density(MUNIC_ESP@data$TASA_PAROW), col="red", lwd=2, main="")
lines(density(MUNIC_ESP@data$TASA_PARO), col="darkgreen", lwd=2)
legend("topright",col=c("red","darkgreen"),legend = c("PAROW","PARO"), lty=1,lwd=2)
```



```

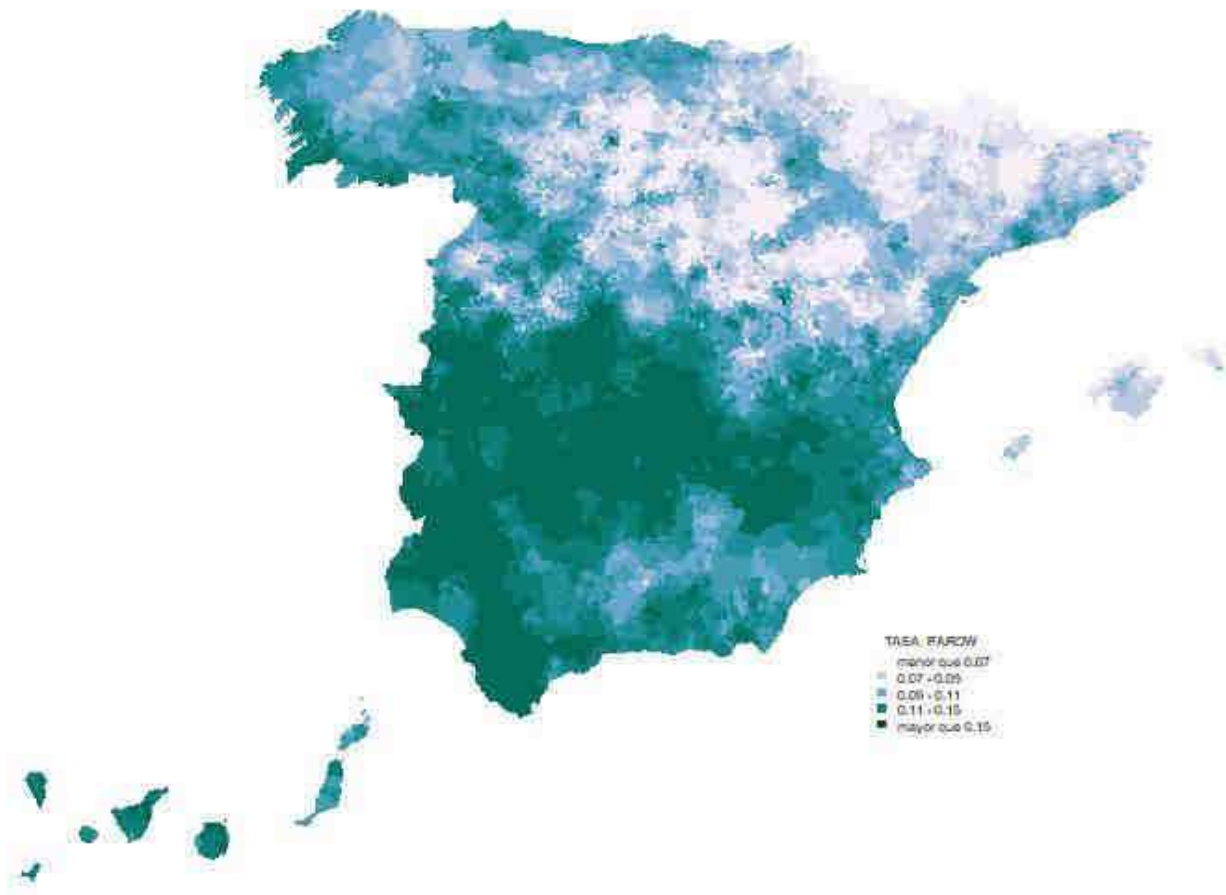
jpeg(file=paste0("graficoTasa de Parow.jpeg"),height=500, width = 500*1.397,bg ="transparent")
ColorBr="PuBuGn"
nombrevbles<-"TASA_PAROW"
par(mar=c(0,0,0,0))
breaks<-classIntervals(eval(parse(text=paste("MUNIC_ESP$",noquote(nombrevbles),sep=""))), 9, style=
  "quantile")
color <- findColours(breaks,palette(brewer.pal(9, ColorBr)))
leg <- findColours(classIntervals(round(eval(parse(text=paste("MUNIC_ESP$",noquote(nombrevbles),sep=
  ""))), digits=2), 5, style="quantile"),
  palette(brewer.pal(5, ColorBr)),under="menor que", over="mayor que", between="-",
  digits=0,cutlabels=FALSE)
#
color <- findColours(breaks,palette(brewer.pal(9, ColorBr)))
leg <- findColours(classIntervals(round(eval(parse(text=paste("MUNIC_ESP$",noquote(nombrevbles),sep=
  ""))), digits=2), 5, style="quantile"),
  palette(brewer.pal(5, ColorBr)),under="menor que", over="mayor que", between="-",
  digits=0,cutlabels=FALSE)
#
plot(MUNIC_ESP,col = color,lty=1, border=NA,lwd=0.25, add=F, axes=F, pch = 19,xlim=c(xmin,xmax), ylim=c(ymin,ymax))
legend(x = -0.75,y = 37,fill=attr(leg, "palette"),
  legend=names(attr(leg,"table")),
  title = nombrevbles, cex=0.6, box.lty=0, border = 0)
dev.off()

```

```

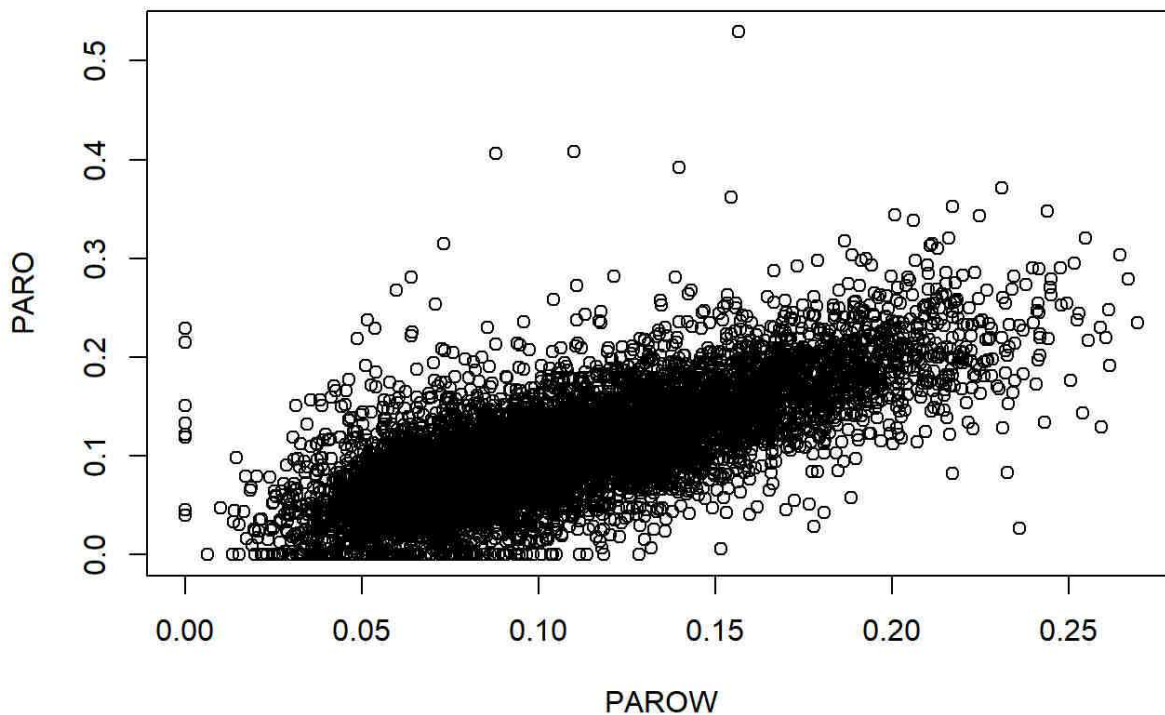
## png
## 2

```



## 6 Correlación entre una variable y su retardo espacial

```
plot(MUNIC_ESP$TASA_PARO~MUNIC_ESP$TASA_PAROW, xlab="PAROW", ylab="PARO")
```



```
# Coeficiente de Correlación  
cor.test(MUNIC_ESP$TASA_PARO,MUNIC_ESP$TASA_PAROW)
```

```
##  
## Pearson's product-moment correlation  
##  
## data: MUNIC_ESP$TASA_PARO and MUNIC_ESP$TASA_PAROW  
## t = 98.964, df = 8106, p-value < 2.2e-16  
## alternative hypothesis: true correlation is not equal to 0  
## 95 percent confidence interval:  
## 0.7296752 0.7493950  
## sample estimates:  
## cor  
## 0.7396939
```

## 6.1 Coeficiente de Correlación espacial: La I de Moran Global

```
moran.test(MUNIC_ESP$TASA_PARO,W,zero.policy = TRUE)
```





## 6.2 Correlación Espacial Local

```
# Correlación Espacial Local
```

```
GLocal<-localG(MUNIC_ESP$TASA_PARO, W, zero.policy = TRUE) # Getis-Ord Statistics  
summary(GLocal)
```

```
##      Min.  1st Qu.   Median     Mean  3rd Qu.    Max.     NA's  
## -5.33117 -1.36245 -0.20896  0.03482  1.15808  9.95585      8
```

```
MUNIC_ESP@data$GLocal<-as.numeric(GLocal)
```

```
# Mapa de la correlación Espacial  
ColorBr="PuBuGn"  
nombrevbles<- "GLocal"  
par(mar=c(0,0,0,0))  
breaks<-classIntervals(eval(parse(text=paste("MUNIC_ESP$",noquote(nombrevbles),sep=""))), 9, style=  
  "quantile",na.rm=TRUE)
```

```
## Warning in classIntervals(eval(parse(text = paste("MUNIC_ESP$"),  
## noquote(nombrevbles), : var has missing values, omitted in finding classes
```

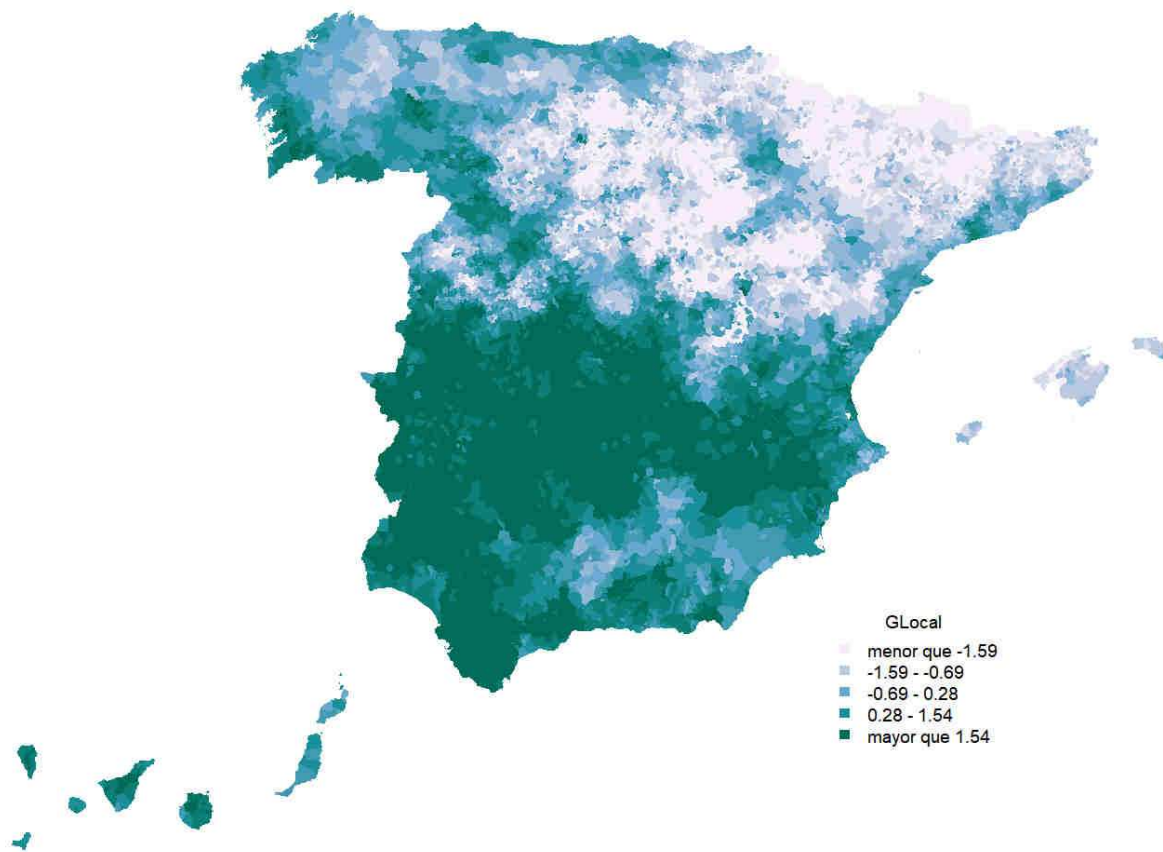
```
color <- findColours(breaks,palette(brewer.pal(5, ColorBr)))  
leg <- findColours(classIntervals(round(eval(parse(text=paste("MUNIC_ESP$",noquote(nombrevbles),sep=  
  "))), digits=2), 5, style="quantile"),  
  palette(brewer.pal(5, ColorBr)),under="menor que", over="mayor que", between="-",  
  digits=0,cutlabels=FALSE)
```

```
## Warning in classIntervals(round(eval(parse(text = paste("MUNIC_ESP$"),  
## noquote(nombrevbles), : var has missing values, omitted in finding classes
```

```
#  
color <- findColours(breaks,palette(brewer.pal(5, ColorBr)))  
leg <- findColours(classIntervals(round(eval(parse(text=paste("MUNIC_ESP$",noquote(nombrevbles),sep=  
  "))), digits=2), 5, style="quantile"),  
  palette(brewer.pal(5, ColorBr)),under="menor que", over="mayor que", between="-",  
  digits=0,cutlabels=FALSE)
```

```
## Warning in classIntervals(round(eval(parse(text = paste("MUNIC_ESP$"),  
## noquote(nombrevbles), : var has missing values, omitted in finding classes
```

```
#  
plot(MUNIC_ESP,col = color,lty=1, border=NA,lwd=0.25, add=F, axes=F, pch = 19,xlim=c(xmin,xmax), ylim=c(ymin,ymax))  
legend(x = -0.75,y = 37,fill=attr(leg, "palette"),  
  legend=names(attr(leg,"table")),  
  title = nombrevbles, cex=0.6, box.lty=0, border = 0)
```



```
par(mar=c(2,2,2,2))
```

```
# Otra forma de estimar la correlación espacial Local
```

```
lmoran<-localmoran(MUNIC_ESP$TASA_PARO,W,zero.policy = TRUE) # I de moran local (para cada poligon  
o)
```

```
summary(lmoran) # la quinta columna proporciona el pValor
```

```
##           Ii           E.Ii           Var.Ii
## Min.      :-3.35429   Min.      :-0.0001234   Min.      :0.0000
## 1st Qu.:  0.01657   1st Qu.: -0.0001234   1st Qu.: 0.1427
## Median :  0.25323   Median : -0.0001234   Median : 0.1665
## Mean      :  0.58320   Mean      :-0.0001232   Mean      : 0.1975
## 3rd Qu.:  0.80532   3rd Qu.: -0.0001234   3rd Qu.: 0.2498
## Max.      :10.51367   Max.      : 0.0000000   Max.      : 0.9996
##
##           Z.Ii           Pr(z > 0)
## Min.      :-7.50337   Min.      :0.0000
## 1st Qu.:  0.04073   1st Qu.: 0.0278
## Median :  0.59216   Median : 0.2769
## Mean      :  1.38678   Mean      : 0.2846
## 3rd Qu.:  1.91418   3rd Qu.: 0.4838
## Max.      :27.68543   Max.      :1.0000
## NA's      :8         NA's      :8
```

```
MUNIC_ESP@data$l Moran<-lmoran[,5]
```

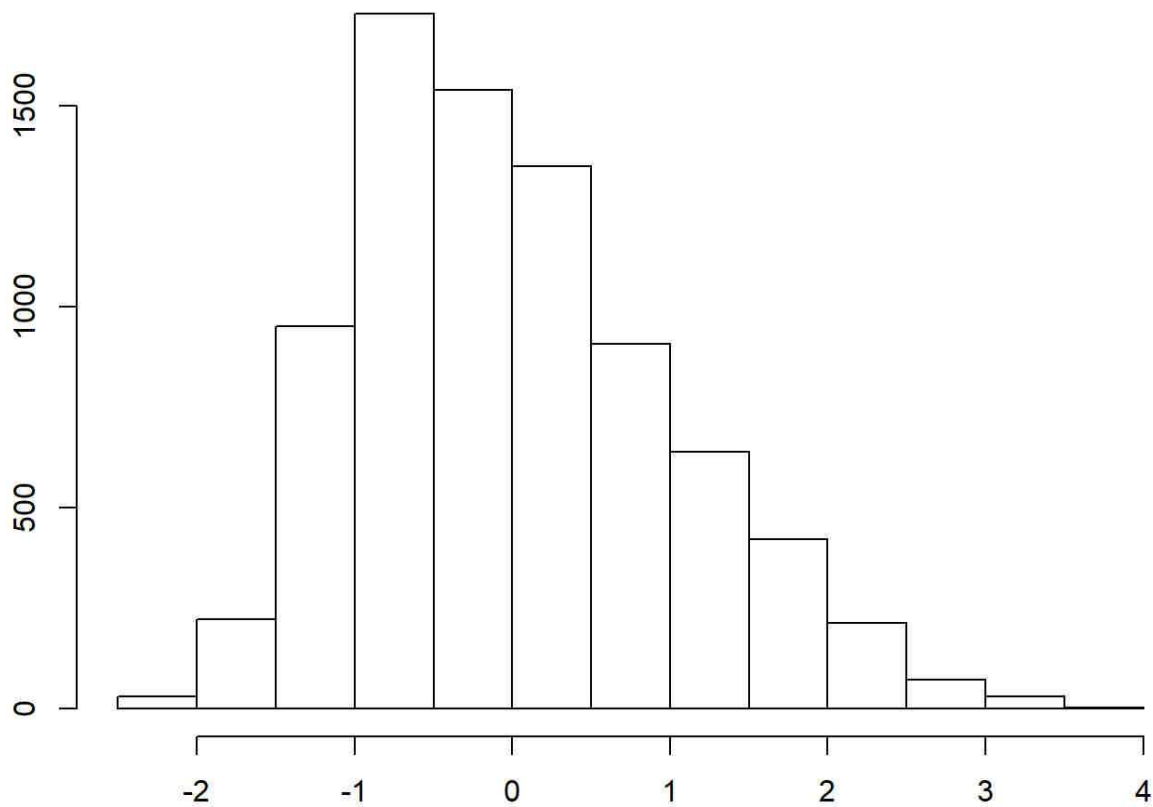
```
MUNIC_ESP@data$quad_sig <- NA
```

```
MUNIC_ESP@data$TASA_PARO_TIP<-scale(MUNIC_ESP@data$TASA_PARO)
```

```
MUNIC_ESP@data$TASA_PAROW_TIP<-scale(MUNIC_ESP@data$TASA_PAROW)
```

```
hist(MUNIC_ESP@data$TASA_PAROW_TIP)
```

**Histogram of MUNIC\_ESP@data\$TASA\_PAROW\_TIP**



### 6.2.1 Mapa de las correlaciones espaciales locales y significativas

```

# high-high quadrant
MUNIC_ESP@data$quad_sig[(MUNIC_ESP@data$TASA_PARO_TIP >= 0 & MUNIC_ESP@data$TASA_PAROW_TIP >= 0 &
  MUNIC_ESP@data$l Moran <= 0.05)] <- "high-high"
# low-low quadrant
MUNIC_ESP@data$quad_sig[(MUNIC_ESP@data$TASA_PARO_TIP <= 0 &
  MUNIC_ESP@data$TASA_PAROW_TIP <= 0 &
  MUNIC_ESP@data$l Moran <= 0.05)] <- "low-low"

# high-low quadrant
MUNIC_ESP@data$quad_sig[(MUNIC_ESP@data$TASA_PARO_TIP >= 0 &
  MUNIC_ESP@data$TASA_PAROW_TIP <= 0 &
  MUNIC_ESP@data$l Moran <= 0.05)] <- "high-low"

# low-high quadrant
MUNIC_ESP@data$quad_sig[(MUNIC_ESP@data$TASA_PARO_TIP <= 0 &
  MUNIC_ESP@data$TASA_PAROW_TIP >= 0 &
  MUNIC_ESP@data$l Moran <= 0.05)] <- "low-high"

# non-significant observations
MUNIC_ESP@data$quad_sig[MUNIC_ESP@data$l Moran > 0.05] <- "not signif."

names(table(MUNIC_ESP@data$quad_sig))

```

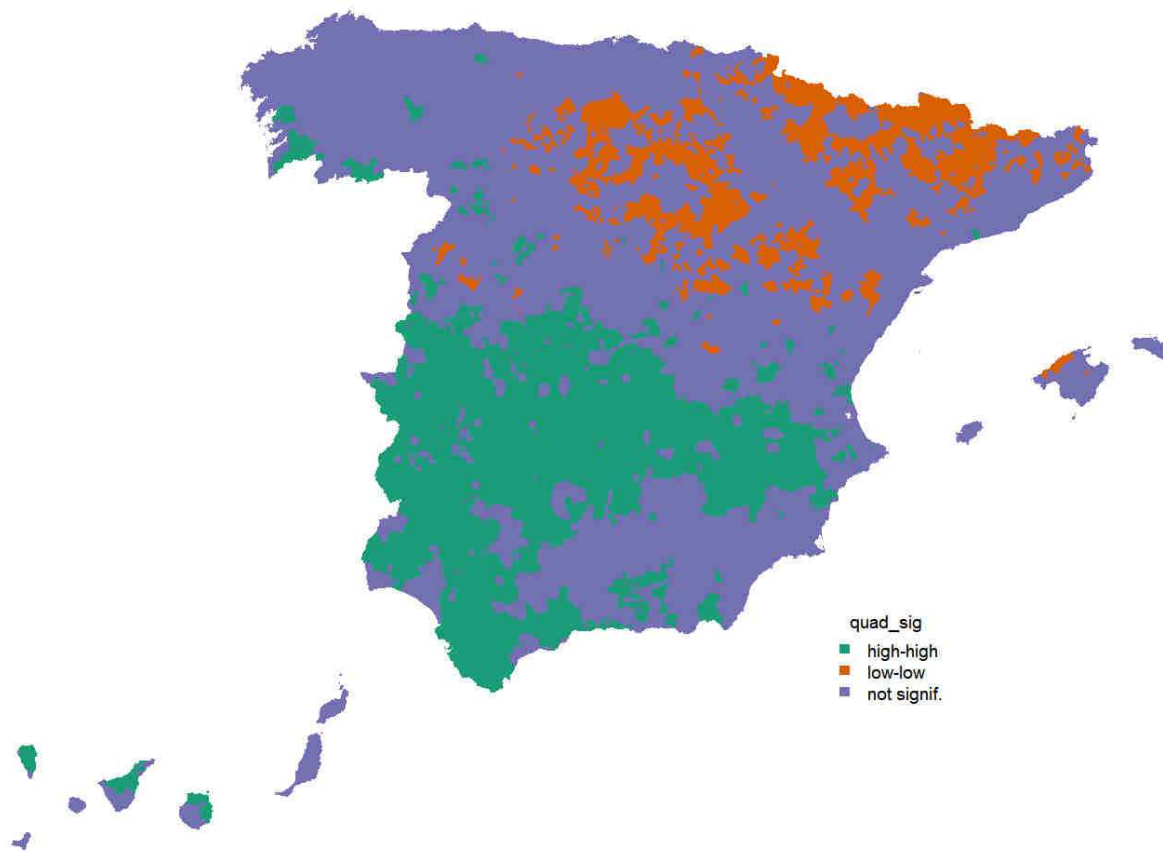
```
## [1] "high-high"    "low-low"      "not signif."
```

```

MUNIC_ESP@data$quad_sig <- as.factor(MUNIC_ESP@data$quad_sig)

nombrevbles<-"quad_sig"
par(mar=c(0,0,0,0))
palette(brewer.pal(n = 5, name = "Dark2"))
plot(MUNIC_ESP,col = eval(parse(text=paste("MUNIC_ESP$",noquote(nombrevbles),sep=""))),
  lty=1, border=NA,lwd=0.25, add=F, axes=F, pch = 19,xlim=c(xmin,xmax), ylim=c(ymin,ymax))
legend(x = -0.75,y = 37, legend=names(table(eval(parse(text=paste("MUNIC_ESP$",noquote(nombrevbles),
  sep=""))))),
  fill=palette(brewer.pal(n = 5, name = "Set2")), title = nombrevbles, cex=0.6, box.lty=0, border = 0)

```



```
par(mar=c(2,2,2,2))
```

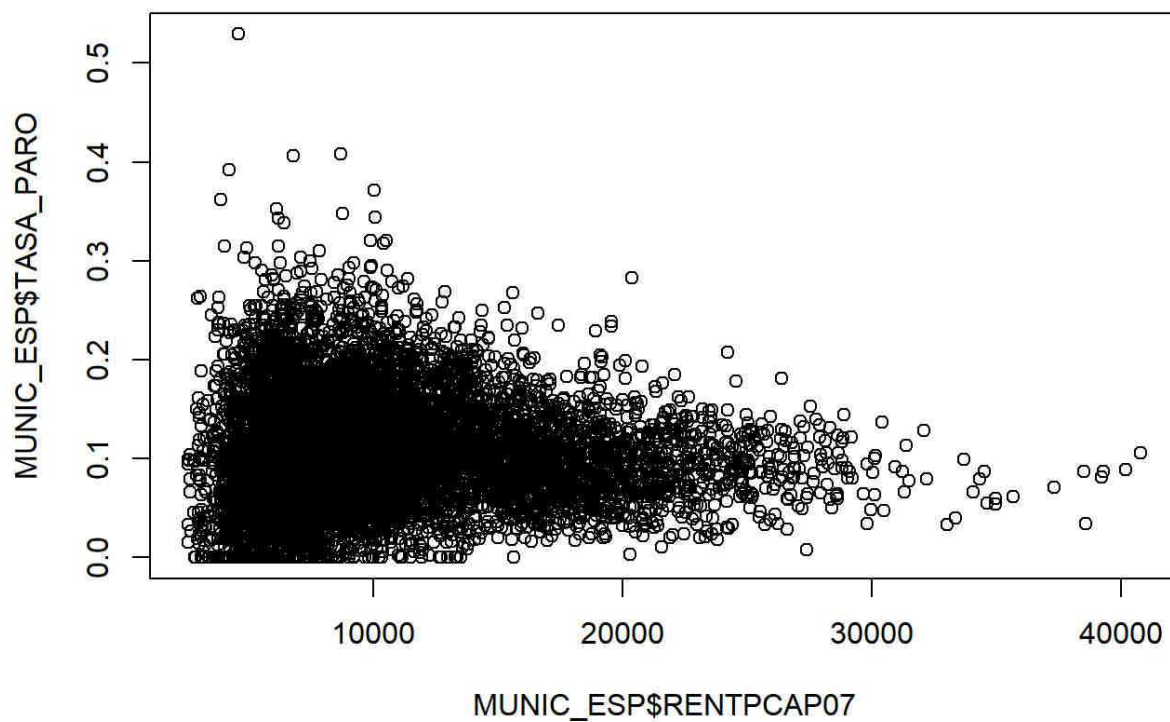
## 7 Correlación entre Tasa de Paro y Reta Percápita

```
# correlación tasa de paro con Rentapercápita
```

```
cor.test(MUNIC_ESP$TASA_PARO,MUNIC_ESP$RENTPCAP07)
```

```
##
## Pearson's product-moment correlation
##
## data: MUNIC_ESP$TASA_PARO and MUNIC_ESP$RENTPCAP07
## t = -4.3957, df = 8106, p-value = 1.118e-05
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
## -0.07045791 -0.02702695
## sample estimates:
## cor
## -0.04876548
```

```
plot(MUNIC_ESP$TASA_PARO~MUNIC_ESP$RENTPCAP07)
```



```
#install.packages("corrplot")
#library(corrplot)
#corrplot(correlacion, method="color")
#correlacion<-cor(MUNIC_ESP@data[,c(58, 9:28)],use = "pairwise.complete")
#correlacion[1,]
```

## 8 Modelo regresion Espacial

### 8.1 Modelo sin efectos espaciales

```
suppressMessages(library(lmtest))

modelo.lm<-lm(TASA_PARO~RENTPCAP07, MUNIC_ESP@data)
summary(modelo.lm)
```

```
##
## Call:
## lm(formula = TASA_PARO ~ RENTPCAP07, data = MUNIC_ESP@data)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.11188 -0.03815 -0.00504  0.03328  0.41830
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  1.134e-01  1.393e-03  81.378 < 2e-16 ***
## RENTPCAP07   -5.173e-07  1.177e-07  -4.396 1.12e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.05567 on 8106 degrees of freedom
## Multiple R-squared:  0.002378,    Adjusted R-squared:  0.002255
## F-statistic: 19.32 on 1 and 8106 DF,  p-value: 1.118e-05
```

```
lmtest::bptest(modelo.lm)# Ho: Ausencia de Heteroscedasticidad
```

```
##
## studentized Breusch-Pagan test
##
## data: modelo.lm
## BP = 234.99, df = 1, p-value < 2.2e-16
```

```
library(sandwich)
modelo.lm.ro<-lmtest::coeftest(modelo.lm, vcov = vcovHC(modelo.lm, "HC0")) # robust; HC0
stargazer(modelo.lm, modelo.lm.ro, type="text")
```

```
##
## =====
##                               Dependent variable:
##                               -----
##                               TASA_PARO
##                               OLS                coefficient
##                               test
##                               (1)                (2)
## -----
## RENTPCAP07                -0.00000***          -0.00000***
##                               (0.00000)          (0.00000)
##
## Constant                   0.113***            0.113***
##                               (0.001)            (0.001)
##
## -----
## Observations                8,108
## R2                          0.002
## Adjusted R2                 0.002
## Residual Std. Error    0.056 (df = 8106)
## F Statistic             19.323*** (df = 1; 8106)
## =====
## Note:                      *p<0.1; **p<0.05; ***p<0.01
```

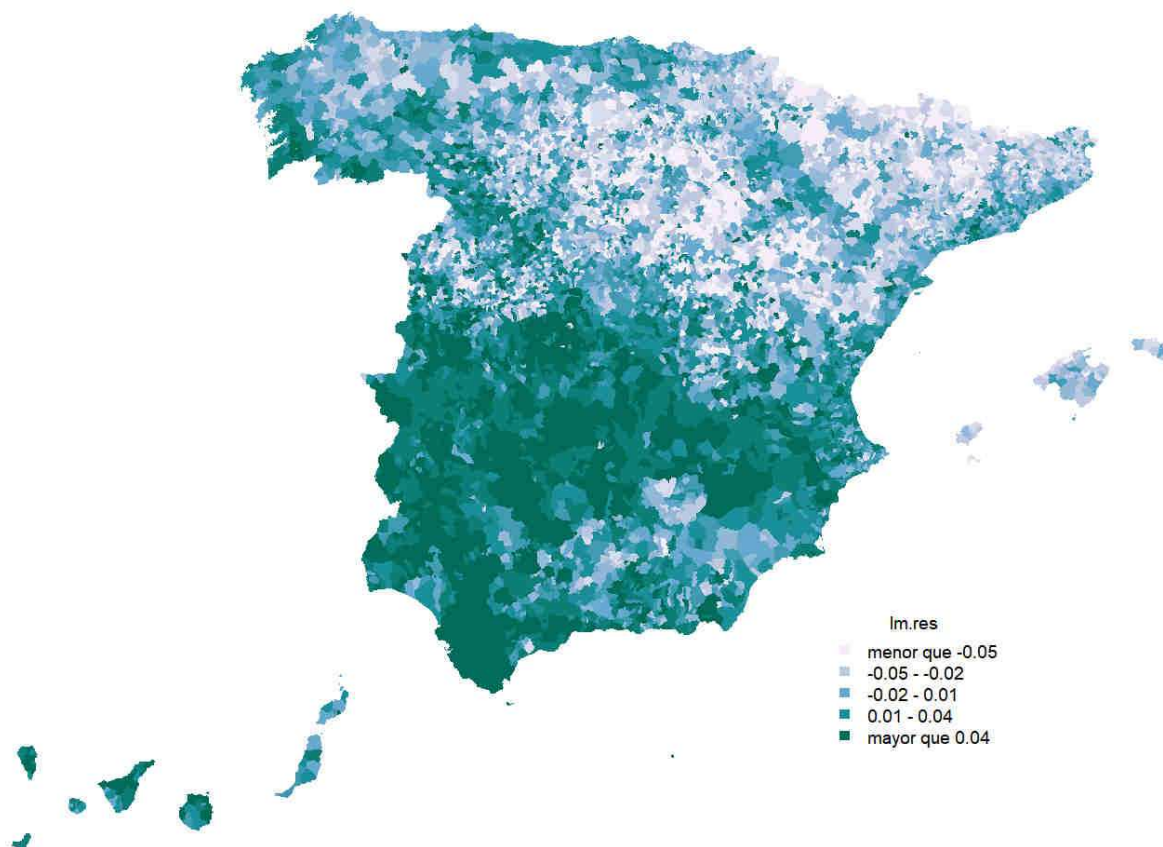
## 8.2 test sobre efectos espaciales en los residuos

```

# Guardo los residuos de cada Municipio
MUNIC_ESP@data$lm.res<-resid(modelo.lm)

ColorBr="PuBuGn"
nombrevbles<-"lm.res"
par(mar=c(0,0,0,0))
breaks<-classIntervals(eval(parse(text=paste("MUNIC_ESP$",noquote(nombrevbles),sep=""))), 9, style=
  "quantile")
color <- findColours(breaks,palette(brewer.pal(5, ColorBr)))
leg <- findColours(classIntervals(round(eval(parse(text=paste("MUNIC_ESP$",noquote(nombrevbles),sep=
  ""))), digits=2), 5, style="quantile"),
  palette(brewer.pal(5, ColorBr)),under="menor que", over="mayor que", between="-",
  digits=0,cutlabels=FALSE)
#
color <- findColours(breaks,palette(brewer.pal(5, ColorBr)))
leg <- findColours(classIntervals(round(eval(parse(text=paste("MUNIC_ESP$",noquote(nombrevbles),sep=
  ""))), digits=2), 5, style="quantile"),
  palette(brewer.pal(5, ColorBr)),under="menor que", over="mayor que", between="-",
  digits=0,cutlabels=FALSE)
#
plot(MUNIC_ESP,col = color,lty=1, border=NA,lwd=0.25, add=F, axes=F, pch = 19,xlim=c(xmin,xmax), ylim=
  c(ymin,ymax))
legend(x = -0.75,y = 37,fill=attr(leg, "palette"),
  legend=names(attr(leg,"table")),
  title = nombrevbles, cex=0.6, box.lty=0, border = 0)

```





```
par(mar=c(2,2,2,2))
```

```
#Si los residuos fuesen un ruido blanco deberían distribuirse de manera uniforme en todo el territorio  
# no debería observarse relación espacial
```

```
# Índice de Moran sobre los residuos  
moran.lm<-lm.morantest(modelo.lm, W, alternative="two.sided", zero.policy=TRUE)  
print(moran.lm)
```

```
##  
## Global Moran I for regression residuals  
##  
## data:  
## model: lm(formula = TASA_PARO ~ RENTPCAP07, data = MUNIC_ESP@data)  
## weights: W  
##  
## Moran I statistic standard deviate = 86.854, p-value < 2.2e-16  
## alternative hypothesis: two.sided  
## sample estimates:  
## Observed Moran I      Expectation      Variance  
##      5.804029e-01      -2.286189e-04      4.469097e-05
```

## 8.3 Test sobre los modelos espaciales

```
# test LM para elegir entre el Spatial LAG (SAR) y el Spatial ERR(SEM)  
LM<-lm.LMtests(modelo.lm, W, test="all", zero.policy=TRUE)  
print(LM)
```

```

##
## Lagrange multiplier diagnostics for spatial dependence
##
## data:
## model: lm(formula = TASA_PARO ~ RENTPCAP07, data = MUNIC_ESP@data)
## weights: W
##
## LMerr = 7543.1, df = 1, p-value < 2.2e-16
##
##
## Lagrange multiplier diagnostics for spatial dependence
##
## data:
## model: lm(formula = TASA_PARO ~ RENTPCAP07, data = MUNIC_ESP@data)
## weights: W
##
## LMlag = 7462.8, df = 1, p-value < 2.2e-16
##
##
## Lagrange multiplier diagnostics for spatial dependence
##
## data:
## model: lm(formula = TASA_PARO ~ RENTPCAP07, data = MUNIC_ESP@data)
## weights: W
##
## RLMerr = 80.273, df = 1, p-value < 2.2e-16
##
##
## Lagrange multiplier diagnostics for spatial dependence
##
## data:
## model: lm(formula = TASA_PARO ~ RENTPCAP07, data = MUNIC_ESP@data)
## weights: W
##
## RLMlag = 0.010354, df = 1, p-value = 0.9189
##
##
## Lagrange multiplier diagnostics for spatial dependence
##
## data:
## model: lm(formula = TASA_PARO ~ RENTPCAP07, data = MUNIC_ESP@data)
## weights: W
##
## SARMA = 7543.1, df = 2, p-value < 2.2e-16

```

## 8.4 Modelos de econometría Espacial

```
formula_lm<-formula(modelo.lm)
```

### 8.4.1 SAR: Spatial lag Model

```
## SAR: Spatial lag Model
```

```
# Se estima por Método Generalizado de los momentos (Ojo que tarda)
```

```
#sar.lm<-lagsarlm(formula_lm, data=MUNIC_ESP@data, W,zero.policy = TRUE)
```

```
#summary(sar.lm)
```

```
#para corregir por heteroscedasticidad habr?a que utilizar la librer?a sphet
```

```
#sar.lm<-sphet::spreg(formula_lm, data=MUNIC_ESP@data, listw=W, model="lag", het=TRUE)
```

```
# o por M?nimos Cuadrados en dos etapas
```

```
sar.ts1m<-stsls(formula_lm, data = MUNIC_ESP@data, W, zero.policy = TRUE,
```

```
na.action = na.fail, robust = TRUE, HC="HC0", legacy=FALSE, W2X = TRUE)
```

```
summary(sar.ts1m)
```

```
##
```

```
## Call:stsls(formula = formula_lm, data = MUNIC_ESP@data, listw = W,
```

```
## zero.policy = TRUE, na.action = na.fail, robust = TRUE, HC = "HC0",
```

```
## legacy = FALSE, W2X = TRUE)
```

```
##
```

```
## Residuals:
```

```
##      Min      1Q    Median      3Q      Max
```

```
## -0.2289284 -0.0229762 -0.0014491  0.0206575  0.3675583
```

```
##
```

```
## Coefficients:
```

```
##      Estimate HC0 std. Error z value Pr(>|z|)
```

```
## Rho      1.1803e+00    1.5255e-01  7.7371 1.021e-14
```

```
## (Intercept) -2.4699e-02    1.7872e-02 -1.3820 0.166981
```

```
## RENTPCAP07  3.6232e-07    1.3941e-07  2.5989 0.009352
```

```
##
```

```
## Residual variance (sigma squared): 0.0015147, (sigma: 0.03892)
```

```
# residuos del Spatial lag model
```

```
sar.res<-resid(sar.ts1m)
```

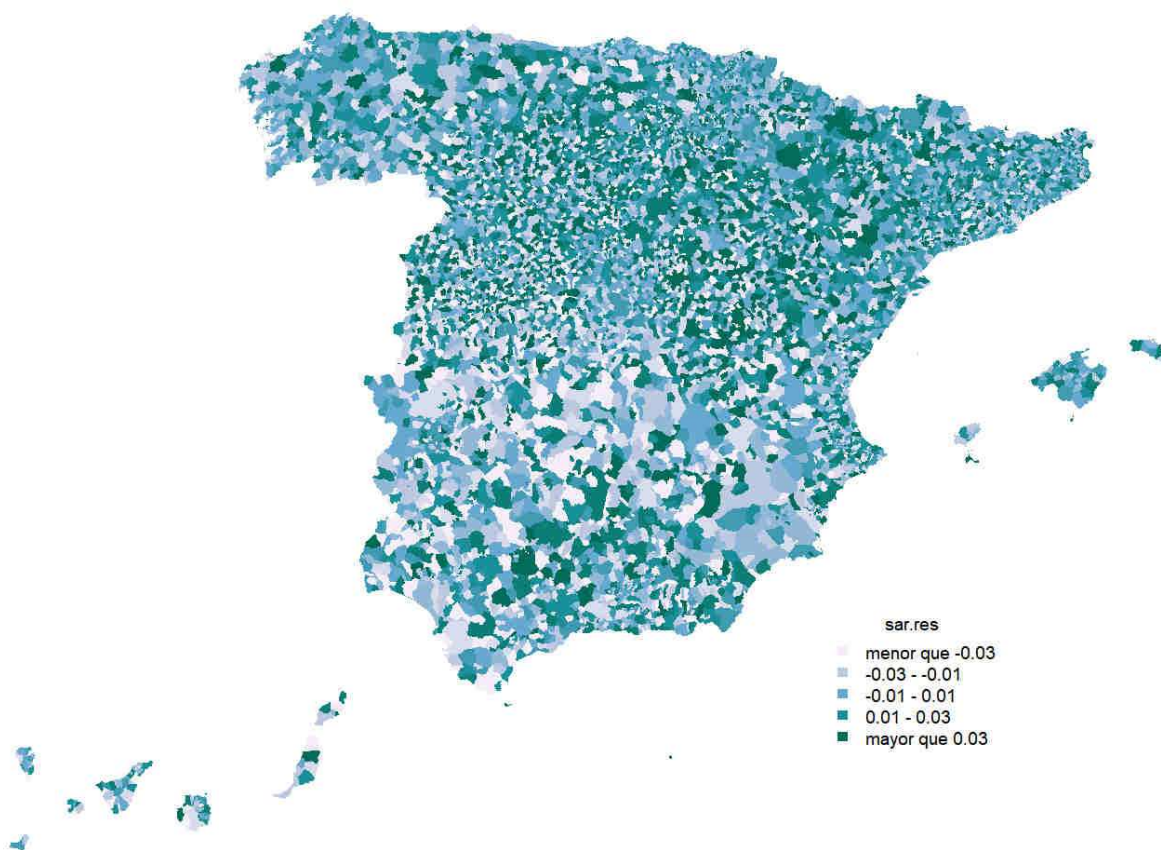
```
names(sar.res)
```

```
## NULL
```

```
# ¿En qué municipios te equivocas más y en qué municipios te equivocas menos?

MUNIC_ESP@data$sar.res<-resid(sar.tslm) #residual PAra poder dibujar los residuos

ColorBr="PuBuGn"
nombrevbles<-"sar.res"
par(mar=c(0,0,0,0))
breaks<-classIntervals(eval(parse(text=paste("MUNIC_ESP$",noquote(nombrevbles),sep=""))), 9, style="quantile")
color <- findColours(breaks,palette(brewer.pal(5, ColorBr)))
leg <- findColours(classIntervals(round(eval(parse(text=paste("MUNIC_ESP$",noquote(nombrevbles),sep="")))), digits=2), 5, style="quantile"),
                  palette(brewer.pal(5, ColorBr)),under="menor que", over="mayor que", between="-",
                  digits=0,cutlabels=FALSE)
#
color <- findColours(breaks,palette(brewer.pal(5, ColorBr)))
leg <- findColours(classIntervals(round(eval(parse(text=paste("MUNIC_ESP$",noquote(nombrevbles),sep="")))), digits=2), 5, style="quantile"),
                  palette(brewer.pal(5, ColorBr)),under="menor que", over="mayor que", between="-",
                  digits=0,cutlabels=FALSE)
#
plot(MUNIC_ESP,col = color,lty=1, border=NA,lwd=0.25, add=F, axes=F, pch = 19,xlim=c(xmin,xmax), ylim=c(ymin,ymax))
legend(x = -0.75,y = 37,fill=attr(leg, "palette"),
      legend=names(attr(leg,"table")),
      title = nombrevbles, cex=0.6, box.lty=0, border = 0)
```

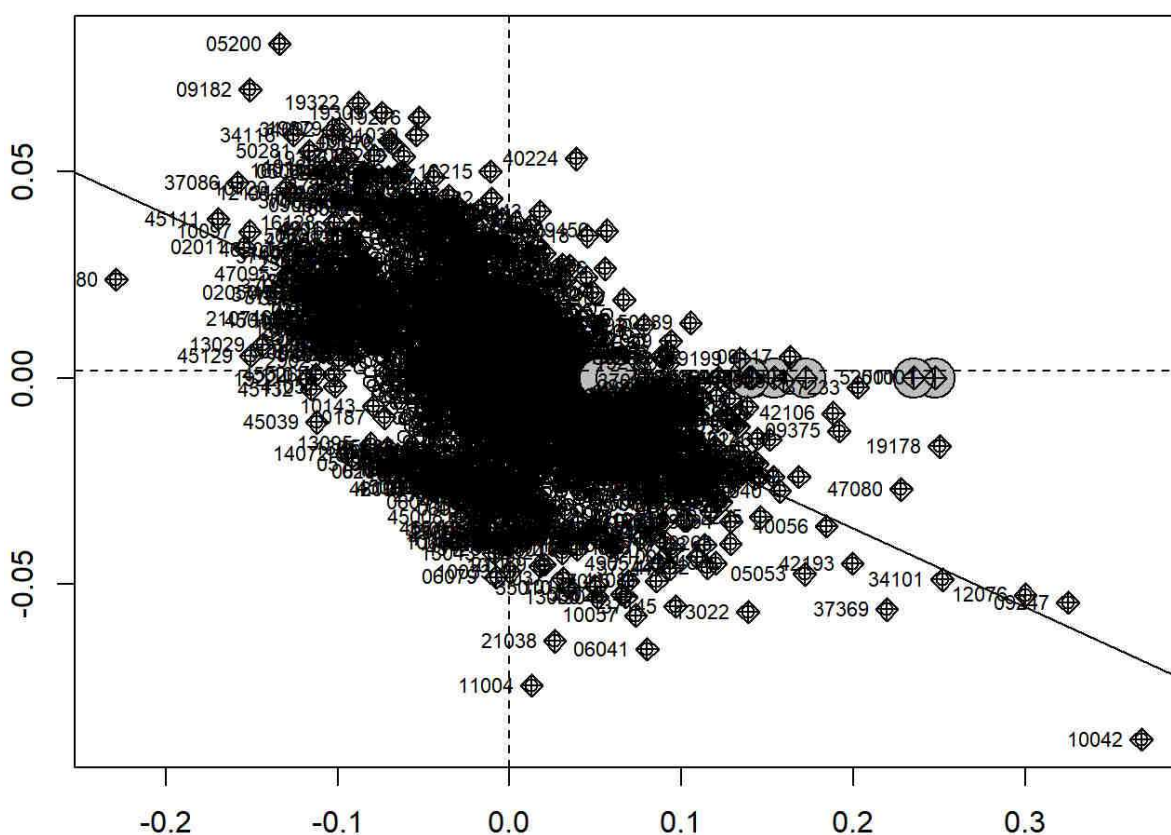


```
par(mar=c(2,2,2,2))
```

```
# Indice de Moran sobre los residuos
moran.test(resid(sar.tslm), W, alternative="greater", zero.policy=TRUE)
```

```
##
## Moran I test under randomisation
##
## data: resid(sar.tslm)
## weights: W n reduced by no-neighbour observations
##
##
## Moran I statistic standard deviate = -28.574, p-value = 1
## alternative hypothesis: greater
## sample estimates:
## Moran I statistic      Expectation      Variance
##      -1.911422e-01      -1.234720e-04      4.468898e-05
```

```
moran.plot(resid(sar.tslm),W,zero.policy = TRUE)
```



```
# Impactos (OJO QUE TARDA MUCHO)
#impacts(sar.lm, listw=W,zero.policy = TRUE )
#impacts(sar.ts1m, listw=W,zero.policy = TRUE )
```

## 8.4.2 SEM: Espatial Error Model

```
## SEM: Espatial Error Model

# por Método Generalizado de los momentos (Ojo que tarda)
#errorsar.lm<-errorsarlm(formula_lm, data=MUNIC_ESP@data, W,zero.policy = TRUE)
#summary(errorsar.lm)

# o corrigiendo por heterocedasticidad
#errorsar.lm<-sphet::spreg(formula_lm, data=MUNIC_ESP@data, listw=W, model="error", het=TRUE)
#summary(errorsar.lm)

# o por Feasible Generalized Least Squares (GLS) with the function GMerrorsar.

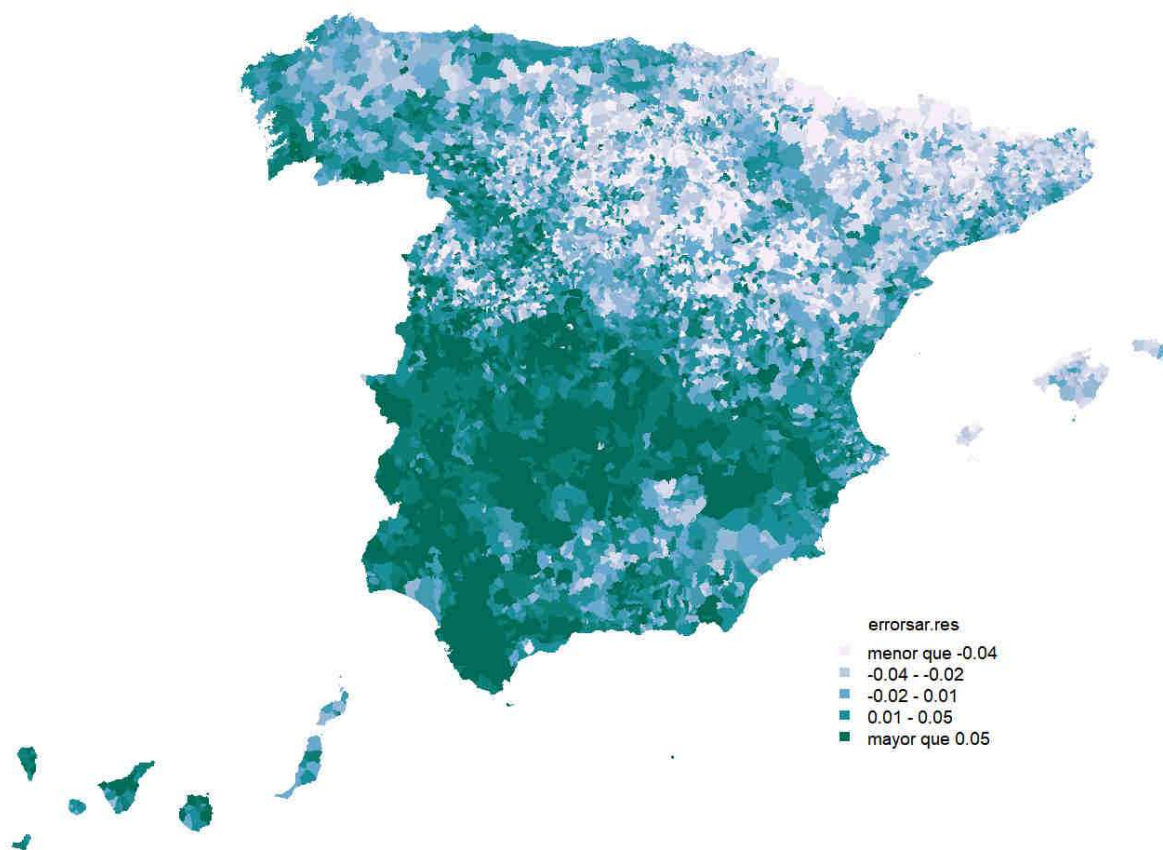
errorsar.fgls<-GMerrorsar(formula_lm, data=MUNIC_ESP@data, W,zero.policy = TRUE)
summary(errorsar.fgls)
```

```
##
## Call:GMerrorsar(formula = formula_lm, data = MUNIC_ESP@data, listw = W,
##      zero.policy = TRUE)
##
## Residuals:
##      Min          1Q      Median          3Q      Max
## -0.1074055 -0.0365009 -0.0030204  0.0359461  0.4266555
##
## Type: GM SAR estimator
## Regions with no neighbours included:
##  07024 17094 36901 46098 46197 46904 51001 52001
## Coefficients: (GM standard errors)
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept) 1.0065e-01 2.2748e-03 44.2448  < 2e-16
## RENTPCAP07  4.3149e-07 1.7273e-07  2.4981  0.01249
##
## Lambda: 0.69563 (standard error): 0.01984 (z-value): 35.062
## Residual variance (sigma squared): 0.0015169, (sigma: 0.038947)
## GM argmin sigma squared: 0.001522
## Number of observations: 8108
## Number of parameters estimated: 4
```

```
# ¿En qué municipios te equivocas más y en qué municipios te equivocas menos?

#MUNIC_ESP@data$errorsar.res<-resid(errorsar.lm) #residual PAra poder dibujar los residuos
MUNIC_ESP@data$errorsar.res<-resid(errorsar.fgls) #residual PAra poder dibujar los residuos

ColorBr="PuBuGn"
nombrevbles<- "errorsar.res"
par(mar=c(0,0,0,0))
breaks<-classIntervals(eval(parse(text=paste("MUNIC_ESP$",noquote(nombrevbles),sep=""))), 9, style="quantile")
color <- findColours(breaks,palette(brewer.pal(5, ColorBr)))
leg <- findColours(classIntervals(round(eval(parse(text=paste("MUNIC_ESP$",noquote(nombrevbles),sep=""))), digits=2), 5, style="quantile"),
                    palette(brewer.pal(5, ColorBr)),under="menor que", over="mayor que", between="-",
                    digits=0,cutlabels=FALSE)
#
color <- findColours(breaks,palette(brewer.pal(5, ColorBr)))
leg <- findColours(classIntervals(round(eval(parse(text=paste("MUNIC_ESP$",noquote(nombrevbles),sep=""))), digits=2), 5, style="quantile"),
                    palette(brewer.pal(5, ColorBr)),under="menor que", over="mayor que", between="-",
                    digits=0,cutlabels=FALSE)
#
plot(MUNIC_ESP,col = color,lty=1, border=NA,lwd=0.25, add=F, axes=F, pch = 19,xlim=c(xmin,xmax), ylim=c(ymin,ymax))
legend(x = -0.75,y = 37,fill=attr(leg, "palette"),
       legend=names(attr(leg,"table")),
       title = nombrevbles, cex=0.6, box.lty=0, border = 0)
```



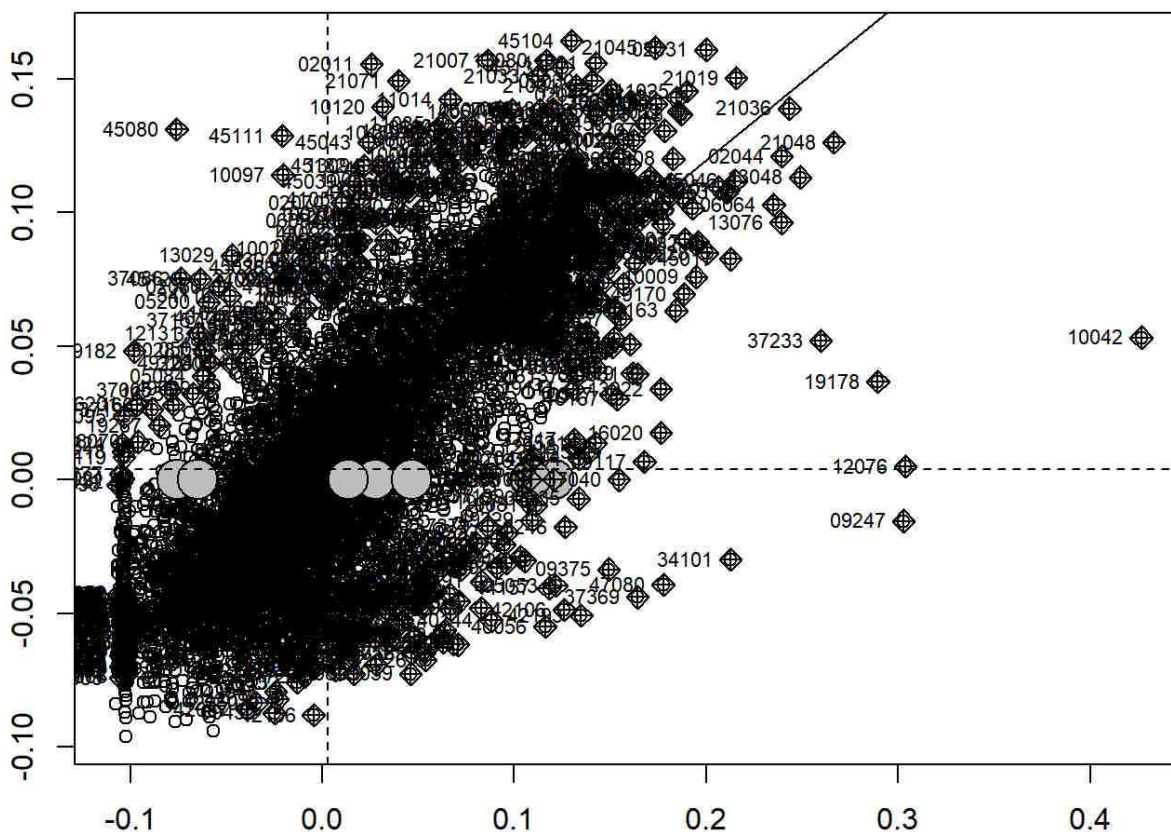


```
par(mar=c(2,2,2,2))
```

```
# Indice de Moran sobre los residuos  
moran.test(resid(errorsar.fgls), W, alternative="greater", zero.policy=TRUE)
```

```
##  
## Moran I test under randomisation  
##  
## data: resid(errorsar.fgls)  
## weights: W n reduced by no-neighbour observations  
##  
##  
## Moran I statistic standard deviate = 87.732, p-value < 2.2e-16  
## alternative hypothesis: greater  
## sample estimates:  
## Moran I statistic      Expectation      Variance  
##      5.864838e-01      -1.234720e-04      4.470714e-05
```

```
moran.plot(resid(errorsar.fgls),W,zero.policy = TRUE)
```





```
##
## Monte-Carlo simulation of Moran I
##
## data: resid(errorsar.fgls)
## weights: W
## number of simulations + 1: 1001
##
## statistic = 0.58648, observed rank = 1001, p-value = 0.000999
## alternative hypothesis: greater
```

```
# RESUMEN de MODELOS
summary(modelo.lm)
```

```
##
## Call:
## lm(formula = TASA_PARO ~ RENTPCAP07, data = MUNIC_ESP@data)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.11188 -0.03815 -0.00504  0.03328  0.41830
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  1.134e-01  1.393e-03  81.378  < 2e-16 ***
## RENTPCAP07   -5.173e-07  1.177e-07  -4.396  1.12e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.05567 on 8106 degrees of freedom
## Multiple R-squared:  0.002378, Adjusted R-squared:  0.002255
## F-statistic: 19.32 on 1 and 8106 DF, p-value: 1.118e-05
```

```
#summary(sar.lm)
summary(sar.tslm)
```

```
##
## Call:stsls(formula = formula_lm, data = MUNIC_ESP@data, listw = W,
##      zero.policy = TRUE, na.action = na.fail, robust = TRUE, HC = "HC0",
##      legacy = FALSE, W2X = TRUE)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.2289284 -0.0229762 -0.0014491  0.0206575  0.3675583
##
## Coefficients:
##              Estimate HC0 std. Error z value Pr(>|z|)
## Rho          1.1803e+00    1.5255e-01  7.7371 1.021e-14
## (Intercept) -2.4699e-02    1.7872e-02 -1.3820  0.166981
## RENTPCAP07   3.6232e-07    1.3941e-07  2.5989  0.009352
##
## Residual variance (sigma squared): 0.0015147, (sigma: 0.03892)
```

```
#summary(errorsar.lm)
summary(errorsar.fgls)
```

```
##
## Call:GMerrorsar(formula = formula_lm, data = MUNIC_ESP@data, listw = W,
##     zero.policy = TRUE)
##
## Residuals:
##      Min        1Q      Median        3Q       Max
## -0.1074055 -0.0365009 -0.0030204  0.0359461  0.4266555
##
## Type: GM SAR estimator
## Regions with no neighbours included:
##  07024 17094 36901 46098 46197 46904 51001 52001
## Coefficients: (GM standard errors)
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept) 1.0065e-01 2.2748e-03 44.2448  < 2e-16
## RENTPCAP07  4.3149e-07 1.7273e-07  2.4981  0.01249
##
## Lambda: 0.69563 (standard error): 0.01984 (z-value): 35.062
## Residual variance (sigma squared): 0.0015169, (sigma: 0.038947)
## GM argmin sigma squared: 0.001522
## Number of observations: 8108
## Number of parameters estimated: 4
```

```
#####
```

## 9 Otros Modelos

```
# Otros modelos Durbin, SARAR

# Durbin está entro del errorsar.lm<-errorsarlm(formula_lm, data=MUNIC_ESP@data, W,zero.policy = TRUE)

# GMM SARAR
#gs2sls <- gsts1s(form, data = data_set,listw = lw)
#sarar_het<-spreg(form, data = data_set,listw = lw, model = "sarar", het=TRUE)

# GMM lag endog
#spreg_lag_endog_het<-spreg(form1, data = data_set, listw = lw, endog = ~police, instruments = ~elect, model = "lag", het=TRUE, lag.instr = TRUE)

# GMM error endog
#spreg_error_endog_het<-spreg(form1, data = data_set, listw = lw, endog = ~police, instruments = ~elect, model = "error", het=TRUE, lag.instr = TRUE)

# GMM sarar endog
#spreg_sarar_endog_het<-spreg(form1, data = data_set, listw = lw, endog = ~police, instruments = ~elect, model = "sarar", het=TRUE, lag.instr = TRUE)

# ML lag mirar library(McSpatial) y el ejemplo en Biban Piras
```



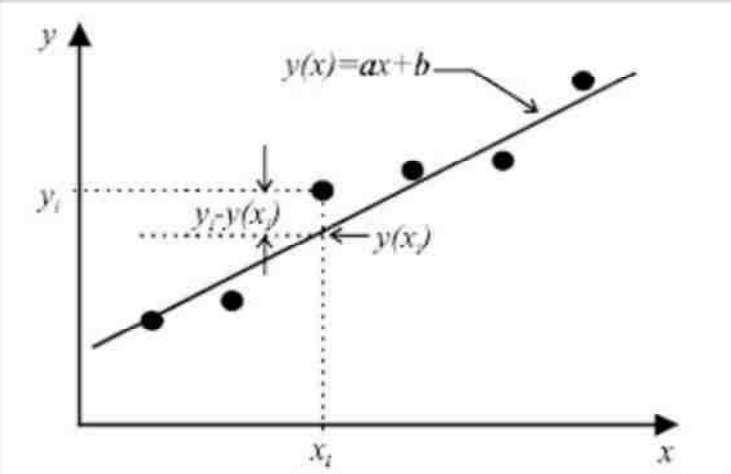
Proyecto de innovación  
 “Innova Docencia” de la UCM  
 Convocatoria 2018-2019  
 Proyecto número 275

## R-adaptación de la asignatura de métodos econométricos en economía y finanzas del Grado de Estadística Aplicada

Facultad de Estudios Estadísticos  
 Departamento de Economía Aplicada, Pública y Política

### Ejemplos de la plataforma Kahoot

Cual de los siguientes supuestos es necesario para que el estimador MCO sea insesgado?



**21**

**0**  
Answers

**Skip**

- ☒ Que el término de error sea Ruido Blanco
- ☐ Que las variables explicativas sean linealmente independientes
- ☐ Que la causalidad vaya en una dirección, sin endogeneidad.
- ☐ Que el término de error se distribuya normal multivariante

Que el término de error sea un ruido Blanco implica que....

24



Skip

0  
Answers



que su autocorrelación sea distinta de cero



Que tenga varianza nula



que se matriz de varianzas y covarianzas sea escalar



que la varianza sea proporcional a su media

sabes que es un "data frame"

23



Skip

0  
Answers



el título del nuevo disco de Jamiroquai



Un objeto para datos de R con todas las columnas numéricas



Es como un objeto list de R pero más general



Un objeto de R, cada columna puede ser de diferente clase

para qué sirve la función library() en R

23



Skip

0

Answers

▲ para acceder directamente a la librería de Amazon


◆ No se sabe, no existe esa función

● para cargar librerías de R

■ Para cargar en memoria datos en R

Qué es RStudio?

15



Skip

0

Answers

▲ Es una versión de R, que mejora los estudios (estadísticos)

◆ Es un interface de usuario más amigable que la consola de R.

● Es la versión de pago de R

■ Es la alternativa a Python y a SAS de software libre

¿Qué es un estimador insesgado?

26



Skip

0  
Answers



Alguien que nos quiere mucho de forma incondicional



Mejor estadístico para aproximar un valor desconocido



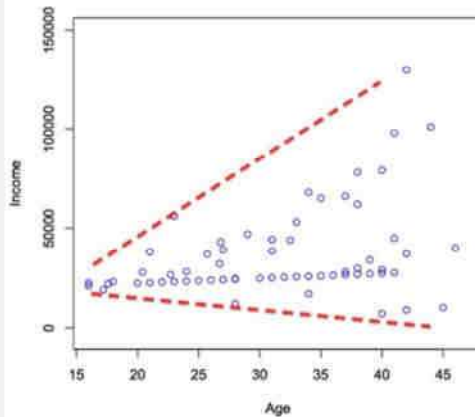
Estadístico sin sesgo positivo para estimar un parámetro



propiedad mínima exigible a un estadístico....apuntar bien

¿Qué es el problema de la Heteroscedasticidad?

22



Skip

0  
Answers



Un problema sin solución



que la varianza de la variable dependiente es muy alta



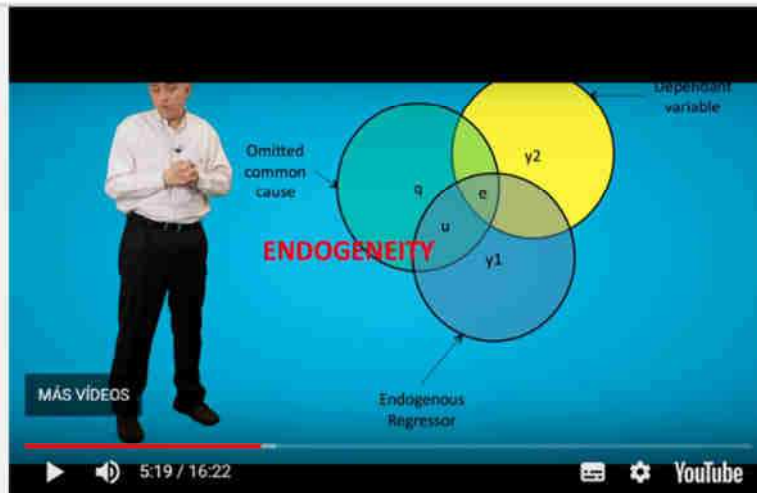
que la varianza del término de error es muy variable



que la varianza de la variable dependiente es muy variable

El problema de la endogeneidad....

30



Skip

0  
Answers



Es que siempre está dentro



Es la principal causa de la obesidad



No impide seguir utilizando MCO



Causa que el Estimador MCO sea Inconsistente